

UNIVERSIDAD AUTONOMA DE MADRID

ESCUELA POLITECNICA SUPERIOR



Grado en Ingeniería Informática

TRABAJO FIN DE GRADO

Detección de perfiles de usuario en MOOCs

Antonio Romo López
Tutora: Rosa M^a Carro Salas

Junio 2018

Detección de perfiles de usuario en MOOCs

AUTOR: Antonio Romo López
TUTORA: Rosa M^a Carro Salas

Grupo GHIA
Escuela Politécnica Superior
Universidad Autónoma de Madrid
Junio de 2018

Resumen (castellano)

Tradicionalmente la enseñanza es presencial, con alumnos y profesores presentes en el aula, esto facilita la creación de una relación profesor-alumno en la que el profesor sabe perfectamente qué alumnos se esfuerzan y cuáles no, sobre qué contenidos necesitan más apoyo y quiénes están en riesgo de no aprobar. Actualmente se están creando plataformas de enseñanza online, que ofrecen cursos llamados MOOCs. Este aprendizaje a distancia no facilita el contacto presencial con los alumnos y el hecho de que un gran número de estudiantes cursen un MOOC y la posibilidad de que cada uno vaya a su propio ritmo, complica realizar un seguimiento del aprendizaje, ofrecer una atención personalizada a cada alumno y apoyarles con los temas en los que presenten mayores dificultades (salvo que lo soliciten explícitamente) o detectar el riesgo de fracaso o abandono de las asignaturas.

Este Trabajo Fin de Grado trata de facilitar el seguimiento de los estudiantes que cursan MOOCs y de predecir los resultados de los mismos (si terminarán el curso de forma exitosa o no), de modo que se puedan tomar las acciones que consideren oportunas según el criterio de los docentes o responsables de cada MOOC. Esto se hace a partir de la información generada a partir de las interacciones de los estudiantes al realizar el MOOC (iniciar un vídeo, realizar un ejercicio, escribir en el foro, etc.), que queda recogida en ficheros de log. Estos logs son utilizados en este trabajo para extraer los atributos con los que predecir el comportamiento y el rendimiento de los alumnos. Los atributos se consiguen de los diferentes eventos que se crean al realizar los MOOCs por parte de los estudiantes.

Tras transformar toda la información disponible en datos útiles, se han utilizado distintos algoritmos de clasificación supervisada y no supervisada (*clustering*) y también una combinación de ambos, con los que se trata de predecir si el alumno superará el curso o no. Se ha buscado cuál de los métodos y algoritmos utilizados aporta unos mejores resultados, y se ha concluido que la combinación de ambos métodos es lo más óptimo.

Palabras clave (castellano)

MOOC, predicción del éxito o fracaso, algoritmos de clasificación supervisada, *clustering*, edX

Abstract (English)

Traditionally the teaching is face-to-face, with students and teachers present in the classroom, this facilitates the creation of a teacher-student relationship in which the teacher knows perfectly which students work and which are not, or in which content they need more support and who are at risk of not pass. Currently, online teaching platforms are being created, offering courses called MOOCs. This distance learning does not facilitate face-to-face contact with students and the fact that a large number of students take a MOOC and the possibility of each one going at their own pace, complicates track the learning, in order to pay personalize attention to each student and support them with the subjects in which they present major difficulties (unless they explicitly request it) or detect the risk of failure or abandonment of the subjects.

This Final Project aims to facilitate the monitoring of students who take MOOCs and to predict their results of them (whether they will finish the course successfully or not), so that they can take the actions they consider appropriate according to the teachers' criteria or responsible for each MOOC. This is done from the information generated from the interactions of the students when performing the MOOC (start a video, perform an exercise, write in the forum, etc.), which is collected in log files. These logs are used in this work to extract the attributes with which to predict the behavior and performance of students. The attributes are obtained from the different events that are created when performing the MOOCs by the students.

After transforming all available information into useful data, different algorithms of supervised and unsupervised classification (clustering) have been used and also a combination of both, with which it is a question of predicting whether the student will pass the course or not. It has been sought which of the methods and algorithms used provides better results, and it has been concluded that the combination of both methods is the best performance.

Keywords (inglés)

MOOCs, prediction of success or failure, supervised classification algorithms clustering, edX

Agradecimientos

Agradezco a Rosa Carro el darme la oportunidad de realizar este trabajo, así como su ayuda en la realización del mismo, sin ella no hubiera sido posible.

A mi familia por su comprensión y apoyo en mis días buenos y malos, no solo durante la realización de este TFG, sino durante toda la carrera.

A todos los compañeros con los que he estado durante la carrera con los que sin duda he pasado grandes momentos.

Y como no puede ser de otra manera a mis amigos, los que siempre están, Lè Groupé. Sin olvidarme de mis amigos del pueblo que me llenan los veranos de alegría y felicidad.

Muchas gracias a todos.

INDICE DE CONTENIDOS

1	Introducción.....	1
1.1.	Motivación.....	1
1.2.	Objetivos.....	1
1.3.	Tecnología utilizada.....	2
1.4.	Organización de la memoria.....	2
2	Estado del arte	3
3	Descripción de los datos.....	5
3.1.	Documentos utilizados.....	5
3.2.	Filtrado de información de eventos	6
4	Tratamiento de los datos.....	7
4.1.	Elección de las Técnicas	7
4.2.	Clasificación	7
4.2.1.	Näive Bayes.....	7
4.2.1.1.	Total.....	8
4.2.1.2.	Semanal	9
4.2.2.	AdaBoost	10
4.2.2.1.	Total.....	10
4.2.2.2.	Semanal	11
4.2.3.	ADTree	14
4.2.3.1.	Total.....	14
4.2.3.2.	Semanal	15
4.2.4.	J48.....	17
4.2.4.1.	Total.....	17
4.2.4.2.	Semanal	18
4.2.5.	Visión global semanal	19
4.2.5.1.	Don Quijote	19
4.2.5.2.	Android.....	21
4.3.	Clustering.....	22
4.3.1.	Don Quijote	22
4.3.1.1.	Total.....	22
4.3.1.2.	Semanal	24
4.3.2.	Android.....	25
4.3.2.1.	Total.....	25
4.3.2.2.	Semanal	25
4.3.3.	Comparación de datos	26
4.4.	Combinada	26
4.4.1.	Don Quijote	27
4.4.2.	Android.....	28
4.4.3.	Comparación de datos	29
5	Conclusiones y trabajo futuro.....	31
5.1.	Conclusiones.....	31
5.2.	Trabajo futuro	31
	Referencias	33
	Glosario	35
	Anexos.....	I
A	Näive Bayes.....	I

A.1	Don Quijote	I
A.2	Android.....	II
B	AdaBoost	V
B.1	Don Quijote	V
B.2	Android.....	VI
C	ADTree	IX
C.1	Don Quijote	IX
C.2	Android.....	X
D	J48.....	XIII
D.1	Don Quijote	XIII
D.2	Android.....	XIV
E	Clusters	XVII
E.1	Don Quijote	XVII
E.2	Android.....	XVIII

INDICE DE FIGURAS

FIGURA 4.1: GRÁFICA EVOLUCIÓN SEMANAL CON ALGORITMO NÄIVE BAYES EN MOOC DON QUIJOTE	9
FIGURA 4.2: GRÁFICA EVOLUCIÓN SEMANAL CON ALGORITMO NÄIVE BAYES EN MOOC ANDROID	10
FIGURA 4.3: GRÁFICA EVOLUCIÓN SEMANAL CON ALGORITMO ADABOOST EN MOOC DON QUIJOTE	12
FIGURA 4.4: GRÁFICA EVOLUCIÓN SEMANAL CON ALGORITMO ADABOOST EN MOOC ANDROID	13
FIGURA 4.5: REPRESENTACIÓN DE LOS NODOS DE DECISIÓN Y PREDICCIÓN DEL ALGORITMO ADTREE	14
FIGURA 4.6: GRAFICA EVOLUCIÓN SEMANAL CON ALGORITMO ADTREE EN MOOC DON QUIJOTE	16
FIGURA 4.7: GRAFICA EVOLUCIÓN SEMANAL CON ALGORITMO ADTREE EN MOOC ANDROID	16
FIGURA 4.8: GRÁFICA EVOLUCIÓN SEMANAL CON ALGORITMO J48 EN MOOC DON QUIJOTE.....	18
FIGURA 4.9: GRÁFICA EVOLUCIÓN SEMANAL CON ALGORITMO J48 EN MOOC ANDROID	19
FIGURA 4.10: GRÁFICA COMPARATIVA PREDICCIÓN DE SUSPENSOS EN MOOC DON QUIJOTE	20
FIGURA 4.11: GRÁFICA COMPARATIVA PREDICCIÓN DE APROBADOS EN MOOC ANDROID	20
FIGURA 4.12: GRÁFICA COMPARATIVA PREDICCIÓN DE SUSPENSOS EN MOOC ANDROID	21
FIGURA 4.13: GRÁFICA COMPARATIVA PREDICCIÓN DE APROBADOS EN MOOC ANDROID	22
FIGURA 4.14: GRÁFICA SEPARACIÓN DE CLÚSTERES PROBLEMAS-EVENTOS MOOC DON QUIJOTE	23
FIGURA 4.15: GRÁFICA CON PORCENTAJE DE ALUMNOS APROBADOS EN CADA CLÚSTER MOOC DON QUIJOTE	24
FIGURA 4.16: GRÁFICA CON PORCENTAJE DE ALUMNOS APROBADOS EN CADA CLÚSTER MOOC ANDROID	26
FIGURA 4.17: GRÁFICA COMPARACIÓN PREDICCIÓN DE SUSPENSOS EN MOOC DON QUIJOTE	27
FIGURA 4.18: GRÁFICA COMPARACIÓN PREDICCIÓN DE APROBADOS EN MOOC DON QUIJOTE.....	28
FIGURA 4.19: GRÁFICA COMPARACIÓN PREDICCIÓN DE SUSPENSOS EN MOOC ANDROID	28
FIGURA 4.20: GRÁFICA COMPARACIÓN PREDICCIÓN DE APROBADOS EN MOOC ANDROID	29

INDICE DE TABLAS

TABLA 4.1: TABLA DE ALUMNOS Y PORCENTAJE DE ALUMNOS APROBADOS Y SUSPENSOS EN MOOC DE DON QUIJOTE CON ALGORITMO NÄIVE BAYES	8
TABLA 4.2: TABLA DE ALUMNOS Y PORCENTAJE DE ALUMNOS APROBADOS Y SUSPENSOS EN MOOC DE ANDROID CON ALGORITMO NÄIVE BAYES	9
TABLA 4.3: TABLA DE ALUMNOS Y PORCENTAJE DE ALUMNOS APROBADOS Y SUSPENSOS EN MOOC DE DON QUIJOTE CON ALGORITMO ADABOOST	11
TABLA 4.4: TABLA DE ALUMNOS Y PORCENTAJE DE ALUMNOS APROBADOS Y SUSPENSOS EN MOOC DE ANDROID CON ALGORITMO ADABOOST.....	11
TABLA 4.5: TABLA DE ALUMNOS SEMANA 1 CON ALGORITMO ADABOOST EN MOOC DON QUIJOTE	12
TABLA 4.6: TABLA DE ALUMNOS SEMANA 3 CON ALGORITMO ADABOOST EN MOOC DON QUIJOTE	12
TABLA 4.7: TABLA DE ALUMNOS SEMANA 3 CON ALGORITMO ADABOOST EN MOOC ANDROID..	13
TABLA 4.8: TABLA DE ALUMNOS SEMANA 4 CON ALGORITMO ADABOOST EN MOOC ANDROID..	13
TABLA 4.9: TABLA DE ALUMNOS Y PORCENTAJE DE ALUMNOS APROBADOS Y SUSPENSOS EN MOOC DE DON QUIJOTE CON ALGORITMO ADTREE	14
TABLA 4.10: TABLA DE ALUMNOS Y PORCENTAJE DE ALUMNOS APROBADOS Y SUSPENSOS EN MOOC DE ANDROID CON ALGORITMO ADTREE	15
TABLA 4.11: TABLA DE ALUMNOS Y PORCENTAJE DE ALUMNOS APROBADOS Y SUSPENSOS EN MOOC DE DON QUIJOTE CON ALGORITMO J48	17
TABLA 4.12: TABLA DE ALUMNOS Y PORCENTAJE DE ALUMNOS APROBADOS Y SUSPENSOS EN MOOC DE ANDROID CON ALGORITMO J48.....	17
TABLA 4.13: TABLA DE ALUMNOS SEMANA 2 CON ALGORITMO J48 EN MOOC DON QUIJOTE	18
TABLA 4.14: TABLA DE ALUMNOS SEMANA 3 CON ALGORITMO J48 EN MOOC DON QUIJOTE	19
TABLA 4.15: TABLA DE LOS ALUMNOS DE CADA CLÚSTER MOOC DON QUIJOTE.....	22
TABLA 4.16: TABLA DE LOS ALUMNOS (PORCENTAJE) DE CADA CLÚSTER MOOC DON QUIJOTE..	23
TABLA 4.17: TABLA DE LOS ALUMNOS DE CADA CLÚSTER MOOC ANDROID.....	25
TABLA 4.18: TABLA DE LOS ALUMNOS (PORCENTAJE) DE CADA CLÚSTER MOOC ANDROID.....	25
TABLA A.1: TABLA DE ALUMNOS SEMANA 1 CON ALGORITMO NÄIVE BAYES EN MOOC DON QUIJOTE	I

TABLA A.2: TABLA DE ALUMNOS SEMANA 2 CON ALGORITMO NÄIVE BAYES EN MOOC DON QUIJOTE	I
TABLA A.3: TABLA DE ALUMNOS SEMANA 3 CON ALGORITMO NÄIVE BAYES EN MOOC DON QUIJOTE	I
TABLA A.4: TABLA DE ALUMNOS SEMANA 4 CON ALGORITMO NÄIVE BAYES EN MOOC DON QUIJOTE	I
TABLA A.5: TABLA DE ALUMNOS SEMANA 5 CON ALGORITMO NÄIVE BAYES EN MOOC DON QUIJOTE	I
TABLA A.6: TABLA DE ALUMNOS SEMANA 6 CON ALGORITMO NÄIVE BAYES EN MOOC DON QUIJOTE	I
TABLA A.7: TABLA DE ALUMNOS SEMANA 7 CON ALGORITMO NÄIVE BAYES EN MOOC DON QUIJOTE	I
TABLA A.8: TABLA DE ALUMNOS SEMANA 8 CON ALGORITMO NÄIVE BAYES EN MOOC DON QUIJOTE	II
TABLA A.9: TABLA DE ALUMNOS SEMANA 9 CON ALGORITMO NÄIVE BAYES EN MOOC DON QUIJOTE	II
TABLA A.10: TABLA DE ALUMNOS SEMANA 10 CON ALGORITMO NÄIVE BAYES EN MOOC DON QUIJOTE	II
TABLA A.11: TABLA DE ALUMNOS SEMANA 1 CON ALGORITMO NÄIVE BAYES EN MOOC ANDROID	II
TABLA A.12: TABLA DE ALUMNOS SEMANA 2 CON ALGORITMO NÄIVE BAYES EN MOOC ANDROID	II
TABLA A.13: TABLA DE ALUMNOS SEMANA 3 CON ALGORITMO NÄIVE BAYES EN MOOC ANDROID	II
TABLA A.14: TABLA DE ALUMNOS SEMANA 4 CON ALGORITMO NÄIVE BAYES EN MOOC ANDROID	II
TABLA A.15: TABLA DE ALUMNOS SEMANA 5 CON ALGORITMO NÄIVE BAYES EN MOOC ANDROID	II
TABLA A.16: TABLA DE ALUMNOS SEMANA 6 CON ALGORITMO NÄIVE BAYES EN MOOC ANDROID	III
TABLA A.17: TABLA DE ALUMNOS SEMANA 7 CON ALGORITMO NÄIVE BAYES EN MOOC ANDROID	III
TABLA A.18: TABLA DE ALUMNOS SEMANA 8 CON ALGORITMO NÄIVE BAYES EN MOOC ANDROID	III
TABLA A.19: TABLA DE ALUMNOS SEMANA 9 CON ALGORITMO NÄIVE BAYES EN MOOC ANDROID	III

TABLA A.20: TABLA DE ALUMNOS SEMANA 10 CON ALGORITMO NÄIVE BAYES EN MOOC ANDROID	III
TABLA B.21: TABLA DE ALUMNOS SEMANA 1 CON ALGORITMO ADABOOST EN MOOC DON QUIJOTE	V
TABLA B.22: TABLA DE ALUMNOS SEMANA 2 CON ALGORITMO ADABOOST EN MOOC DON QUIJOTE	V
TABLA B.23: TABLA DE ALUMNOS SEMANA 3 CON ALGORITMO ADABOOST EN MOOC DON QUIJOTE	V
TABLA B.24: TABLA DE ALUMNOS SEMANA 4 CON ALGORITMO ADABOOST EN MOOC DON QUIJOTE	V
TABLA B.25: TABLA DE ALUMNOS SEMANA 5 CON ALGORITMO ADABOOST EN MOOC DON QUIJOTE	V
TABLA B.26: TABLA DE ALUMNOS SEMANA 6 CON ALGORITMO ADABOOST EN MOOC DON QUIJOTE	V
TABLA B.27: TABLA DE ALUMNOS SEMANA 7 CON ALGORITMO ADABOOST EN MOOC DON QUIJOTE	V
TABLA B.28: TABLA DE ALUMNOS SEMANA 8 CON ALGORITMO ADABOOST EN MOOC DON QUIJOTE	V
TABLA B.29: TABLA DE ALUMNOS SEMANA 9 CON ALGORITMO ADABOOST EN MOOC DON QUIJOTE	VI
TABLA B.30: TABLA DE ALUMNOS SEMANA 10 CON ALGORITMO ADABOOST EN MOOC DON QUIJOTE	VI
TABLA B.31: TABLA DE ALUMNOS SEMANA 1 CON ALGORITMO ADABOOST EN MOOC ANDROID	VI
TABLA B.32: TABLA DE ALUMNOS SEMANA 2 CON ALGORITMO ADABOOST EN MOOC ANDROID	VI
TABLA B.33: TABLA DE ALUMNOS SEMANA 3 CON ALGORITMO ADABOOST EN MOOC ANDROID	VI
TABLA B.34: TABLA DE ALUMNOS SEMANA 4 CON ALGORITMO ADABOOST EN MOOC ANDROID	VI
TABLA B.35: TABLA DE ALUMNOS SEMANA 5 CON ALGORITMO ADABOOST EN MOOC ANDROID	VI
TABLA B.36: TABLA DE ALUMNOS SEMANA 6 CON ALGORITMO ADABOOST EN MOOC ANDROID	VI
TABLA B.37: TABLA DE ALUMNOS SEMANA 7 CON ALGORITMO ADABOOST EN MOOC ANDROID	VII
TABLA B.38: TABLA DE ALUMNOS SEMANA 8 CON ALGORITMO ADABOOST EN MOOC ANDROID	VII
TABLA B.39: TABLA DE ALUMNOS SEMANA 9 CON ALGORITMO ADABOOST EN MOOC ANDROID	VII

TABLA B.40: TABLA DE ALUMNOS SEMANA 10 CON ALGORITMO ADABOOST EN MOOC ANDROID	VII
TABLA C.41: TABLA DE ALUMNOS SEMANA 1 CON ALGORITMO ADTREE EN MOOC DON QUIJOTE	IX
TABLA C.42: TABLA DE ALUMNOS SEMANA 2 CON ALGORITMO ADTREE EN MOOC DON QUIJOTE	IX
TABLA C.43: TABLA DE ALUMNOS SEMANA 3 CON ALGORITMO ADTREE EN MOOC DON QUIJOTE	IX
TABLA C.44: TABLA DE ALUMNOS SEMANA 4 CON ALGORITMO ADTREE EN MOOC DON QUIJOTE	IX
TABLA C.45: TABLA DE ALUMNOS SEMANA 5 CON ALGORITMO ADTREE EN MOOC DON QUIJOTE	IX
TABLA C.46: TABLA DE ALUMNOS SEMANA 6 CON ALGORITMO ADTREE EN MOOC DON QUIJOTE	IX
TABLA C.47: TABLA DE ALUMNOS SEMANA 7 CON ALGORITMO ADTREE EN MOOC DON QUIJOTE	IX
TABLA C.48: TABLA DE ALUMNOS SEMANA 8 CON ALGORITMO ADTREE EN MOOC DON QUIJOTE	IX
TABLA C.49: TABLA DE ALUMNOS SEMANA 9 CON ALGORITMO ADTREE EN MOOC DON QUIJOTE	X
TABLA C.50: TABLA DE ALUMNOS SEMANA 10 CON ALGORITMO ADTREE EN MOOC DON QUIJOTE	X
TABLA C.51: TABLA DE ALUMNOS SEMANA 1 CON ALGORITMO ADTREE EN MOOC ANDROID.....	X
TABLA C.52: TABLA DE ALUMNOS SEMANA 2 CON ALGORITMO ADTREE EN MOOC ANDROID.....	X
TABLA C.53: TABLA DE ALUMNOS SEMANA 3 CON ALGORITMO ADTREE EN MOOC ANDROID.....	X
TABLA C.54: TABLA DE ALUMNOS SEMANA 4 CON ALGORITMO ADTREE EN MOOC ANDROID.....	X
TABLA C.55: TABLA DE ALUMNOS SEMANA 5 CON ALGORITMO ADTREE EN MOOC ANDROID.....	X
TABLA C.56: TABLA DE ALUMNOS SEMANA 6 CON ALGORITMO ADTREE EN MOOC ANDROID.....	X
TABLA C.57: TABLA DE ALUMNOS SEMANA 7 CON ALGORITMO ADTREE EN MOOC ANDROID....	XI
TABLA C.58: TABLA DE ALUMNOS SEMANA 8 CON ALGORITMO ADTREE EN MOOC ANDROID....	XI
TABLA C.59: TABLA DE ALUMNOS SEMANA 9 CON ALGORITMO ADTREE EN MOOC ANDROID....	XI
TABLA C.60: TABLA DE ALUMNOS SEMANA 10 CON ALGORITMO ADTREE EN MOOC ANDROID..	XI

TABLA D.61: TABLA DE ALUMNOS SEMANA 1 CON ALGORITMO J48 EN MOOC DON QUIJOTE ..	XIII
TABLA D.62: TABLA DE ALUMNOS SEMANA 2 CON ALGORITMO J48 EN MOOC DON QUIJOTE ..	XIII
TABLA D.63: TABLA DE ALUMNOS SEMANA 3 CON ALGORITMO J48 EN MOOC DON QUIJOTE ..	XIII
TABLA D.64: TABLA DE ALUMNOS SEMANA 4 CON ALGORITMO J48 EN MOOC DON QUIJOTE ..	XIII
TABLA D.65: TABLA DE ALUMNOS SEMANA 5 CON ALGORITMO J48 EN MOOC DON QUIJOTE ..	XIII
TABLA D.66: TABLA DE ALUMNOS SEMANA 6 CON ALGORITMO J48 EN MOOC DON QUIJOTE ..	XIII
TABLA D.67: TABLA DE ALUMNOS SEMANA 7 CON ALGORITMO J48 EN MOOC DON QUIJOTE ..	XIII
TABLA D.68: TABLA DE ALUMNOS SEMANA 8 CON ALGORITMO J48 EN MOOC DON QUIJOTE ..	XIII
TABLA D.69: TABLA DE ALUMNOS SEMANA 9 CON ALGORITMO J48 EN MOOC DON QUIJOTE ..	XIV
TABLA D.70: TABLA DE ALUMNOS SEMANA 10 CON ALGORITMO J48 EN MOOC DON QUIJOTE	XIV
TABLA D.71: TABLA DE ALUMNOS SEMANA 1 CON ALGORITMO J48 EN MOOC ANDROID	XIV
TABLA D.72: TABLA DE ALUMNOS SEMANA 2 CON ALGORITMO J48 EN MOOC ANDROID	XIV
TABLA D.73: TABLA DE ALUMNOS SEMANA 3 CON ALGORITMO J48 EN MOOC ANDROID	XIV
TABLA D.74: TABLA DE ALUMNOS SEMANA 4 CON ALGORITMO J48 EN MOOC ANDROID	XIV
TABLA D.75: TABLA DE ALUMNOS SEMANA 5 CON ALGORITMO J48 EN MOOC ANDROID	XIV
TABLA D.76: TABLA DE ALUMNOS SEMANA 6 CON ALGORITMO J48 EN MOOC ANDROID	XIV
TABLA D.77: TABLA DE ALUMNOS SEMANA 7 CON ALGORITMO J48 EN MOOC ANDROID	XV
TABLA D.78: TABLA DE ALUMNOS SEMANA 8 CON ALGORITMO J48 EN MOOC ANDROID	XV
TABLA D.79: TABLA DE ALUMNOS SEMANA 9 CON ALGORITMO J48 EN MOOC ANDROID	XV
TABLA D.80: TABLA DE ALUMNOS SEMANA 10 CON ALGORITMO J48 EN MOOC ANDROID	XV
TABLA E.81: TABLA DE ALUMNOS SEMANA 1 SEPARADOS POR CLÚSTERES EN MOOC DON QUIJOTE	XVII
TABLA E.82: TABLA DE ALUMNOS SEMANA 2 SEPARADOS POR CLÚSTERES EN MOOC DON QUIJOTE	XVII
TABLA E.83: TABLA DE ALUMNOS SEMANA 3 SEPARADOS POR CLÚSTERES EN MOOC DON QUIJOTE	XVII
TABLA E.84: TABLA DE ALUMNOS SEMANA 4 SEPARADOS POR CLÚSTERES EN MOOC DON QUIJOTE	XVII

TABLA E.85: TABLA DE ALUMNOS SEMANA 5 SEPARADOS POR CLÚSTERES EN MOOC DON QUIJOTE	XVII
TABLA E.86: TABLA DE ALUMNOS SEMANA 6 SEPARADOS POR CLÚSTERES EN MOOC DON QUIJOTE	XVII
TABLA E.87: TABLA DE ALUMNOS SEMANA 7 SEPARADOS POR CLÚSTERES EN MOOC DON QUIJOTE	XVII
TABLA E.88: TABLA DE ALUMNOS SEMANA 8 SEPARADOS POR CLÚSTERES EN MOOC DON QUIJOTE	XVII
TABLA E.89: TABLA DE ALUMNOS SEMANA 9 SEPARADOS POR CLÚSTERES EN MOOC DON QUIJOTE	XVIII
TABLA E.90: TABLA DE ALUMNOS SEMANA 10 SEPARADOS POR CLÚSTERES EN MOOC DON QUIJOTE	XVIII
TABLA E.91: TABLA DE ALUMNOS SEMANA 1 SEPARADOS POR CLÚSTERES EN MOOC ANDROID	XVIII
TABLA E.92: TABLA DE ALUMNOS SEMANA 2 SEPARADOS POR CLÚSTERES EN MOOC ANDROID	XVIII
TABLA E.93: TABLA DE ALUMNOS SEMANA 3 SEPARADOS POR CLÚSTERES EN MOOC ANDROID	XVIII
TABLA E.94: TABLA DE ALUMNOS SEMANA 4 SEPARADOS POR CLÚSTERES EN MOOC ANDROID	XVIII
TABLA E.95: TABLA DE ALUMNOS SEMANA 5 SEPARADOS POR CLÚSTERES EN MOOC ANDROID	XVIII
TABLA E.196: TABLA DE ALUMNOS SEMANA 6 SEPARADOS POR CLÚSTERES EN MOOC ANDROID	XVIII
TABLA E.97: TABLA DE ALUMNOS SEMANA 7 SEPARADOS POR CLÚSTERES EN MOOC ANDROID	XIX
TABLA E.98: TABLA DE ALUMNOS SEMANA 8 SEPARADOS POR CLÚSTERES EN MOOC ANDROID	XIX
TABLA E.99: TABLA DE ALUMNOS SEMANA 9 SEPARADOS POR CLÚSTERES EN MOOC ANDROID	XIX
TABLA E.100: TABLA DE ALUMNOS SEMANA 10 SEPARADOS POR CLÚSTERES EN MOOC ANDROID	XIX

INDICE DE FÓRMULAS

FÓRMULA 4.1: FÓRMULA TEOREMA DE BAYES.....	7
FÓRMULA 4.2: FÓRMULA DE DISTRIBUCIÓN GAUSSIANA.....	8

1 Introducción

1.1. Motivación

El ser humano es curioso por naturaleza, por ello solemos tener interés por aprender cosas nuevas. Desde pequeños empezamos en el colegio y empezamos a desarrollar esa curiosidad y buscando las soluciones a los problemas propuestos. De mayores continuamos apuntándonos a cursos, de inglés, de música...etc. En la actualidad, cada vez es más fácil obtener conocimiento a través de Internet, buscando información que hace referencia a cualquier campo de aprendizaje. Esto ha llevado a la creación de cursos gratuitos de aprendizaje, MOOCs (*Massive Open On-line Course*). Estos cursos son ofertados por numerosas universidades, y facilitan a las personas sin recursos necesarios para poder acceder a cursos que conllevan una inversión económica [10]0.

Debido al carácter gratuito de los cursos, el número de abandono de estos cursos es muy alto. En otros casos, el usuario no tiene la necesidad de obtener este certificado y con el conocimiento adquirido se da por satisfecho. Además, estos cursos no tienen un límite de tiempo ni obligan al usuario a ir al día, por lo que el control del profesor sobre los alumnos es prácticamente inexistente. Realizar un seguimiento de las acciones de los alumnos en los MOOCs es una tarea muy complicada. Más aún lo es tratar de predecir, a la vista de las acciones de un usuario, si está interesado realmente en el curso o no y, en caso de estarlo, si va por buen camino o existe riesgo de fracaso o de abandono. La motivación de este TFG es proporcionar las herramientas necesarias para poder realizar un seguimiento de los estudiantes y, en particular, para detectar perfiles de estudiantes que, aun estando verdaderamente interesados en el curso, presentan riesgo de fracaso. Para ello se utilizarán diferentes algoritmos de clasificación y de *clustering*, para tratar de poner en conocimiento del profesor cuáles de los alumnos sí están interesados y en cuáles de ellos debe centrar más su atención.

1.2. Objetivos

Como ya se ha mencionado, el abandono de este tipo de cursos es muy alto. El principal objetivo de este TFG es tratar de dar al profesor de MOOCs información sobre el estado de sus estudiantes. Debido a esta alta tasa de abandono, sería interesante dar a conocer al profesor los alumnos en los que debe centrarse por considerarse “rescatables”, e incluso poder animar o avisar a aquellos estudiantes que están en riesgo de fracasar en las evaluaciones de que no se olviden de preparar y realizar los ejercicios. El profesor está mucho más limitado en estos cursos online que en cursos presenciales a la hora de ofrecer retroalimentación a los estudiantes. Con esto se trata de reducir esas limitaciones, pudiendo dar un trato más personalizado al estudiante.

Lógicamente una vez dada esta información sobre los estudiantes al profesor, le corresponde a él decidir el uso que hace de ella. Pero no solamente es útil para el profesor. El propio estudiante puede recibir información sobre la posición en la que se encuentra, si es suficiente el trabajo que está desarrollando o si por el contrario debería dedicar más tiempo al curso para acabarlo de forma satisfactoria.

Para conseguir dar retroalimentación al profesor sobre el desempeño y el posible futuro de los estudiantes, se estudiarán diferentes métodos. Uno de los objetivos a alcanzar en este trabajo consiste precisamente en vislumbrar cuál de estos métodos proporciona una información mayor y más fiable al profesor.

Como objetivo secundario se tratará de obtener los atributos que proporcionen la información más relevante (de entre los disponibles o calculables a partir de los logs de los MOOCs) para facilitar la clasificación de los alumnos con los distintos algoritmos de predicción.

1.3. Tecnología utilizada

En este apartado se van a enumerar las tecnologías utilizadas:

Python [4]: Lenguaje de programación [3] utilizado para el desarrollo de programas con los que extraer los datos de los logs.

Gedit [5]: Editor de texto de propósito general, utilizado para la creación de código y visualización de los resultados.

Weka [6]: Programa de software enfocado en el aprendizaje automático y minería de datos, fabricado por la universidad de Waikato. Utilizado para la realización de predicciones con diferentes algoritmos.

Microsoft Word [7]: Procesador de texto utilizado para realizar esta memoria.

EasyBib [8]: Gestor de referencias.

Gnuplot [9]: Generador de gráficas.

Shotwell Viewer [17]: Editor de imágenes

1.4. Organización de la memoria

La memoria consta de los siguientes capítulos:

- El capítulo 1 se expone la motivación que da lugar a la realización de este TFG, así como los objetivos a cumplir para dar por satisfactorio el trabajo y las herramientas utilizadas. Además, se detallará la estructura de este documento.
- En el capítulo 2 se presenta el estado del arte. En él se dará una explicación sobre que son los MOOCs y cuál ha sido su evolución a lo largo del tiempo. También se hablará de otros trabajos o artículos relacionados con el tema y las conclusiones que han sacado.
- En el capítulo 3 se describen los datos utilizados en este trabajo. Para ello se explicará que documentos han sido utilizados y el filtrado que se ha llevado a cabo para conseguir los atributos necesarios para el desarrollo de este trabajo.
- En el capítulo 4 se explicarán las técnicas utilizadas y los resultados obtenidos tras utilizarlas. Además, se comparan los resultados obtenidos de cada una de las técnicas discutiendo cuál de ellas es la óptima para conseguir los objetivos del TFG.
- En el capítulo 5 se exponen las conclusiones del TFG y el trabajo que se puede realizar en un futuro.
- Finalmente, este documento termina con las referencias bibliográficas utilizadas y varios anexos, en los que se incluyen los datos con los resultados obtenidos semanalmente con cada uno de los algoritmos.

2 Estado del arte

MOOC es el acrónimo en inglés de *Masive Online Open Course*. En español su acrónimo es COMA Curso On-line Masivo y Abierto. Como su propio nombre indica, estos cursos son impartidos por Internet con la intención de llegar al mayor número de personas posible, un número ilimitado. Además, posee la característica de ser abierto: todo el mundo puede acceder a ellos sin necesidad de pertenecer a la institución que lo imparte y sin la necesidad de pagar por ello, al menos en el acceso al contenido. Es posible que estos cursos cobren por la obtención del título o por tutorías con los profesores.

La falta de necesidad de acudir a un aula da total libertad geográfica para cursar el curso. Facilita el autoaprendizaje proporcionando todos los contenidos a través de la red como pueden ser documentos y vídeos además de proporcionar lugares de debate, foros.

En agosto de 2007 se creó el primer MOOC conceptual en la universidad de Utah, cuyo autor fue David Wiley. Tras este curso otras universidades continuaron con esta iniciativa. El crecimiento de los MOOCs fue progresivo llegando a 160.000 personas en un MOOC de inteligencia artificial de la universidad de Stanford en el año 2011. Los MOOCs ya eran una realidad, de tal forma que el periódico New York Times publica en el 2012 un artículo titulado "El año de los MOOCs". Actualmente más de 800 universidades de todo el mundo ofrecen estos cursos.

Existen plataformas que alojan todo tipo de MOOCs. Algunas de estas son INTEF, Miriada X, Udacity, Udemy o Coursera. En este TFG los datos analizados pertenecen a la plataforma edX.

Con respecto al abandono en MOOCs, en [14] se afirma que la forma de combatir el abandono en los MOOCs se basa en adelantarse al alumno. En el artículo se afirma que es posible conocer que un alumno va a abandonar el curso una semana antes, consiguiendo un 70 % de acierto en sus predicciones. Con este conocimiento se puede tratar de evitar este abandono motivándole para que continúe. Para conseguir esto han creado modelos con datos de alumnos previos que utilizaran en tiempo real para preverlo. A esto lo llaman *transfer learning*.

Se comentan principalmente dos factores a la hora de abandonar los MOOCs. El primero es que no todo el mundo se inscribe en estos cursos con la intención de acabarlo. La segunda causa es que existen motivos personales como el trabajo u otras obligaciones que les impiden terminarlo.

Con respecto a la predicción de las actuaciones de los estudiantes, en el trabajo descrito en [13] se realiza un estudio en el cual se trata de predecir si un alumno de un MOOC realizará o no un ejercicio. Si se considera que sí lo va a realizar, decidirá si lo resolverá de manera satisfactoria o no. Además, se trata de predecir la calificación que obtendrá el alumno en el examen final del curso. Para conseguir estos objetivos se basan en los ejercicios realizados por parte de los alumnos durante el curso.

Durante el estudio de un curso MOOC, cada atributo tiene una correlación diferente con los alumnos que han aprobado, y por ello unos son más importantes que otros a la hora de escoger cuáles utilizar para predecir qué alumnos van por el buen camino y cuáles no. En el TFM descrito en [15] se explica que hay atributos (como, por ejemplo, el sistema operativo utilizado) que no influyen en la predicción, mientras que otros como el número de aciertos y el número de problemas realizados sobresalen, con una correlación superior al 90% con la

clase aprobado. Otros atributos con una gran correlación según este trabajo serían el número de eventos (75%), el número de vídeos (72%) y el número de sesiones (65%).

En el trabajo anterior, los algoritmos de clasificación predicen con un mayor porcentaje de acierto los alumnos que aprobarán [15]. En él se comparan algoritmos de clasificación con *Long Short Term Memory* que es una red neuronal recurrente. En él se admite que es probable que con un número mayor de variables es posible que se obtenga un mejor rendimiento con algoritmos de *Deep learning* que con el número de variables que se han utilizado para este estudio, con el que no se aprecia ninguna mejora en la clasificación con respecto a algoritmos tradicionales de clasificación. El algoritmo que proporciona los mejores resultados en este caso se basa en árboles de decisión.

Con respecto a la construcción de perfiles de usuarios, en [16] se habla de obtener perfiles a través de transiciones de eventos consecutivos. Para ello utilizan los logs del MOOC "Jugando con Android – Aprende a programar tu primera aplicación". Estos eventos serían: actividades (respuesta a una pregunta sobre una actividad de programación prevista para la semana en la que se encuentra), ejercicios (respuesta a una pregunta asociada a un vídeo), vídeos (ver vídeo), proyecto (entrega de parte del proyecto de programación) foro (acción en el foro) y documentos (leer o descargar documento de texto relacionado con un vídeo). Tras estudiar las transiciones que realizan los alumnos se han creado seis clústeres y las conclusiones son:

- Los alumnos cuya mayoría de transiciones son de documento a documento (D2D) son personas que no van a aprobar el curso.
- Los alumnos que basan sus interacciones en eventos de vídeos y ejercicios (V2V, V2E, E2V y E2E) pero que solamente actúan al inicio durante un breve periodo de tiempo son aquellos estudiantes que empiezan el curso, pero pronto lo abandonan.
- Los alumnos que basan sus actividades en los vídeos (V2V) y que tienen picos menores en ejercicios y documento a documento y además están activos hasta el final del curso sin hacer actividades evaluables son aquellos que están interesados en el curso, pero no tienen la intención de sacarse un certificado.
- Los alumnos que realizan algunos ejercicios y actividades y finalmente descargan los documentos abandonando la actividad en el curso bastante pronto son aquellos que no han podido continuar con el curso, pero descargan los documentos para revisarlos más tarde.
- Los estudiantes que hacen ejercicios (E2E) y actividades (A2A) pero no tocan el contenido obtienen buenos resultados
- Los estudiantes que hacen ejercicios y actividades y además tienen interacciones con el contenido son los estudiantes normales (mayor número de usuarios) y los que sacan los mejores resultados.

En este trabajo se va a explorar el uso de diferentes algoritmos de clasificación, algoritmos de *clustering* y la combinación de ambos tipos de algoritmos con el objetivo descrito anteriormente. Además, se van a realizar las pruebas en dos MOOCs diferentes, comparando los resultados obtenidos en ambos para precisar las conclusiones.

3 Descripción de los datos

Los datos utilizados en este trabajo han sido proporcionados por la Universidad Autónoma de Madrid. En ellos se encuentran datos correspondientes a MOOCs alojados en la plataforma edX [12], más concretamente a los MOOCs de La España del Quijote y Jugando con Android – Aprende a programar tu primera App, todos ellos anonimizados.

3.1. Documentos utilizados

Los archivos utilizados para la realización de este TFG correspondientes a cada uno de los cursos han sido los siguientes:

❖ La España del Quijote:

- CertificateQuijote501x1T2015.csv: Este archivo contiene información de las notas de los usuarios que han realizado el MOOC de la España del Quijote. De él se han extraído los usuarios que han conseguido el certificado y los que por el contrario no lo han conseguido, pero lo más importante es el número con el que identificar a los usuarios, que es la primera piedra para poder llevar a cabo el análisis de sus interacciones. En próximas secciones del documento se hablará de este documento como “certificado”
- SocialQuijote501x1T2015.csv: Este archivo contiene información sobre el foro del MOOC de la España del Quijote. De él se ha extraído el número de comentarios de cada uno de los usuarios. En próximas secciones del documento se hablará de este documento como “social”
- Registro de eventos: Son una serie de archivos que contienen todos los eventos del MOOC La España del Quijote, separados por cada día transcurrido en el que se han producido eventos. Siguen el siguiente formato “1T2015anonymized_Quijote501x-uamx-edx-events-AAAA-MM-DD-edx.txt” en el que AAAA corresponde al año, MM al mes y DD al día en el que ocurren dichos eventos. Son 729 archivos que comprenden del 29 de septiembre de 2014 al 14 de febrero de 2017. De ellos obtenemos los eventos que posteriormente serán explicados.

❖ Jugando con Android – Aprende a programar tu primera App:

- CertificateAndroid301x1T2015.csv: Este archivo contiene información de las notas de los usuarios que han realizado el MOOC de Jugando con Android – Aprende a programar tu primera App. Cumple el mismo objetivo que CertificateQuijote501x1T2015.csv en la España del Quijote.
- SocialAndroid301x1T2015.csv: Este archivo contiene información sobre el foro del MOOC Jugando con Android – Aprende a programar tu primera App y cumple el mismo objetivo que SocialQuijote501x1T2015.csv en la España del Quijote.
- Registro de eventos: Son una serie de archivos que contienen todos los eventos del MOOC Jugando con Android – Aprende a programar tu primera App. Siguen el siguiente formato “1T2015anonymized_Android501x-uamx-edx-events-AAAA-MM-DD-edx.txt” Son 818 archivos que van del 16 de noviembre de 2014 al 9 de febrero de 2017. Cumple el mismo objetivo que el registro de eventos de la España del Quijote.

3.2. Filtrado de información de eventos

En esta sección se describen principalmente los eventos recogidos de los documentos llamados en la sección 4.1 registro de eventos. Para que nos sean de utilidad se debe filtrar su contenido, para así poder asociar unos atributos a cada usuario de los MOOCs. Antes de esto hay que entender que estos documentos son un log de las acciones acometidas por los usuarios, y que cada acción tiene unos parámetros diferentes a través de los cuales podremos diferenciar que tipo de evento ha realizado el usuario. Para empezar, es necesario conocer cuáles son los eventos en los que estamos interesados de cara al estudio posterior y una vez elegidos estos eventos es necesario saber el formato en el que se van a encontrar. Los tipos de eventos existentes, así como su formato o representación en el fichero de log, se pueden conseguir en la “Guía de Iniciación a los datos de EdX” [11] y en el manual de EdX [1].

Una vez conocidos los formatos de cada evento, se ha procedido a realizar la carga de atributos de cada usuario. Para ello se han creado dos scripts en bash que adaptan el formato y escogen los eventos adecuados y los atributos que son de interés. Una vez ejecutados sendos scripts, entra en juego un programa realizado en Python, encargado del filtrado de los eventos, y se encargará de calcular una serie de atributos que serán utilizados para la realización de la detección de perfiles de usuario en MOOCs. Además de esto se procederá a la clasificación de los alumnos en aprobados o en suspensos. A continuación, se enumerarán y comentarán cada uno de estos atributos utilizados.

Para empezar, vamos a hablar de los atributos conseguidos en los archivos certificado y social. De certificado obtenemos el id de cada uno de los participantes del curso y si ha obtenido certificado con la nota. De social obtenemos el número de comentarios. Gracias al atributo id podremos relacionar los eventos realizados por cada usuario. Sin tener conocimiento de cuál es el motivo, se ha comprobado que hay algunos eventos realizados por usuarios no inscritos dentro de los MOOCs; dichos eventos serán descartados si no coinciden con un usuario inscrito en el documento certificado. El hecho de si un estudiante obtiene o no el certificado será utilizado como atributo evaluador de clase, es decir, es el atributo que diferencia unos estudiantes de otros en la clasificación.

Pese a que cada tipo de evento tiene su formato, todos los tipos de eventos tienen en común algunos parámetros, como por ejemplo la identificación del usuario, el evento generado, el tipo de dicho evento y el instante en el que se produce. Debido a esto se ha podido calcular el número total de todos los eventos generados por cada usuario.

Los eventos de tipo problemas y exámenes han proporcionado información sobre los aciertos y los errores cometidos por cada estudiante, obteniendo de ellos los atributos aciertos e intentos. Además, se conseguirá el número real de problemas realizados, ya que un problema solo contará la primera vez que sea realizado.

Los eventos de tipo vídeos han proporcionado información sobre los vídeos visitados, así como los movimientos hacia delante y hacia atrás en ellos. Se obtienen así tres nuevos atributos: el número de vídeos vistos (solo cuenta una vez cada vídeo, aunque lo haya visto más de una vez), el número de movimientos hacia atrás en el vídeo (vuelve a ver alguna parte del vídeo), y el número de movimientos hacia adelante (se salta parte del vídeo).

Por último, se utilizará el número de sesiones por cada usuario, obtenido a partir de los distintos eventos.

Finalmente, se enumeran los atributos utilizados: id, certificado, nota, comentarios, eventos, aciertos, intentos, problemas, vídeos, movAtras, movDelante y sesion.

4 Tratamiento de los datos

4.1. Elección de las Técnicas

En este apartado se van a explicar los resultados obtenidos de someter los datos a diferentes métodos de clasificación, *clustering* y combinación de estos dos. Antes de continuar, cabe mencionar la diferencia principal entre clasificación y *clustering*. Las técnicas de clasificación son de aprendizaje supervisado. Esto quiere decir que, a partir de un conjunto de elementos ya clasificados y etiquetados conforme a esa clasificación, nuevos elementos serán clasificados en los grupos predeterminados. Las técnicas de *clustering*, por su parte son de aprendizaje no supervisado: todos los elementos parten sin etiquetar y se dividen en grupos según las semejanzas entre ellos; en este caso los grupos no existen previamente, sino que se forman según dichas semejanzas. Además de explorar la utilización de ambas técnicas por separado y combinadas, utilizando el conjunto total de datos disponibles, también se va a analizar qué ocurre cuando se trabaja con los datos agrupados semanalmente. La utilidad de saber si aprueba o suspende un alumno con todos los datos disponibles hasta la finalización o abandono del curso no es realmente el objetivo, ya que con eso el profesor no consigue saber en qué alumnos debe concentrar sus esfuerzos durante el proceso de aprendizaje. Sin embargo, sí nos proporcionará una visión sobre el porcentaje de acierto máximo que podemos esperar en el análisis realizado cada semana, y una primera visión de cuál de los algoritmos de clasificación tiene un mejor rendimiento. El objetivo realmente se consigue en las predicciones semanales, que nos darán una idea, cada semana, sobre qué alumnos van por el buen camino para aprobar el MOOC, y así saber en qué alumnos concentrar sus fuerzas, y en cuales simplemente entender que no tienen una motivación real para realizar este curso.

4.2. Clasificación

En este apartado se describe cómo se clasifican los alumnos en aprobados o suspensos en función de los eventos recogidos anteriormente. Para esto se ha utilizado el programa Weka [6] y como algoritmos de clasificación se han escogido Näive Bayes, AdaBoost, ADTree y J48.

4.2.1. Näive Bayes

El algoritmo Näive Bayes se considera un clasificador bayesiano ingenuo. Este algoritmo toma como base el teorema de Bayes, que calcula la probabilidad de que suceda un evento A en función de que haya ocurrido anteriormente un evento B. El teorema de Bayes tiene una enorme relevancia debido a que vincula la probabilidad de A dado B con la probabilidad de B dado A, se expresa con la fórmula de la figura 4.1:

$$p(A|B) = \frac{p(A)p(B|A)}{p(B)}$$

Fórmula 4.1: Fórmula teorema de Bayes

Donde:

- $P(A|B)$ es la probabilidad de A dado B.
- $P(A)$ probabilidad a priori de que suceda A.

- $P(B|A)$ probabilidad de B dado A.
- $P(B)$ probabilidad a priori de que suceda B.

Este algoritmo es llamado N  ive o ingenuo debido a que sus caracter  sticas son totalmente independientes entre s  , o lo que es lo mismo, que se d   una determinada caracter  stica o no, no afecta a la presencia o ausencia del resto de las caracter  sticas. Para este clasificador, se ha asumido que los atributos siguen distribuciones gaussianas, utilizando la f  rmula 4.2:

$$P(x = v|c) = \frac{1}{\sqrt{2\pi\sigma_c^2}} e^{-\frac{(v-\mu_c)^2}{2\sigma_c^2}}$$

F  rmula 4.2: F  rmula de distribuci  n gaussiana

Donde:

- μ_c es la media del atributo asociado a la clase c
- σ_c^2 la varianza del atributo asociada a la clase c

4.2.1.1. Total

4.2.1.1.1. Don Quijote

En esta secci  n se muestran los resultados de aplicar el algoritmo N  ive Bayes al total de los datos obtenidos del MOOC La Espa  a de Don Quijote, clasificando los estudiantes en aprobados y suspendidos. En la siguiente tabla se pueden observar los resultados.

<i>Alumnos</i>	<i>Suspense</i>	<i>Aprobado</i>
<i>Suspense</i>	1708	56
<i>Aprobado</i>	1	198
<i>Porcentaje</i>	<i>Suspense</i>	<i>Aprobado</i>
<i>Suspense</i>	99.9	22
<i>Aprobado</i>	0.1	78

Tabla 4.1: Tabla de alumnos y porcentaje de alumnos aprobados y suspensos en MOOC de Don Quijote con algoritmo N  ive Bayes

Como se puede observar en la tabla 4.1, los alumnos que se considera que suspenden pr  cticamente suspenden (99,9%). Sin embargo, de los alumnos que se considera que van a aprobar un 22% suspenden. Es una cantidad significativa, por lo que podemos decir que este no es un buen algoritmo para la clasificaci  n de los estudiantes.

4.2.1.1.2. Android

En este apartado se muestran los resultados de aplicar el algoritmo N  ive Bayes al total de los datos obtenidos del MOOC Jugando con Android – Aprende a programar tu primera App, clasificando los estudiantes en aprobados y suspendidos. En la siguiente tabla se pueden observar los resultados.

<i>Alumnos</i>	<i>Suspense</i>	<i>Aprobado</i>
<i>Suspense</i>	7836	207
<i>Aprobado</i>	2	491
<i>Porcentaje</i>	<i>Suspense</i>	<i>Aprobado</i>
<i>Suspense</i>	100	29.7
<i>Aprobado</i>	0	70.3

Tabla 4.2: Tabla de alumnos y porcentaje de alumnos aprobados y suspensos en MOOC de Android con algoritmo N  ive Bayes

Como ocurr  a en la tabla 4.1, en la tabla 4.2 podemos afirmar que pr  cticamente se cumple la previsi  n en los casos en los que a los alumnos se les clasifica como suspensos, pero tiene el mismo problema, hasta un 29.7% de los alumnos que se considera que aprueban en realidad, suspenden. Por este motivo se confirma que este algoritmo no es   ptimo para la clasificaci  n de los estudiantes.

4.2.1.2. Semanal

4.2.1.2.1. Don Quijote

En este apartado se va a exponer la evoluci  n semanal de aplicar el algoritmo de N  ive Bayes en el MOOC La Espa  a de Don Quijote. Como se puede ver en la figura 4.1, la predicci  n mejora con el paso de las semanas, acerc  ndose al 78% de acierto en usuarios clasificados como aprobados que ten  amos en la clasificaci  n con todos los datos. Se observa claramente c  mo la mejora de predicci  n se corta al llegar a la semana 9. La predicci  n de suspensos es muy buena en porcentaje desde la semana 1, pero hay que tener en cuenta que la gran mayor  a de los usuarios registrados suspenden. Una vez m  s, mostramos semanalmente que no es un algoritmo   ptimo para la clasificaci  n de los alumnos.

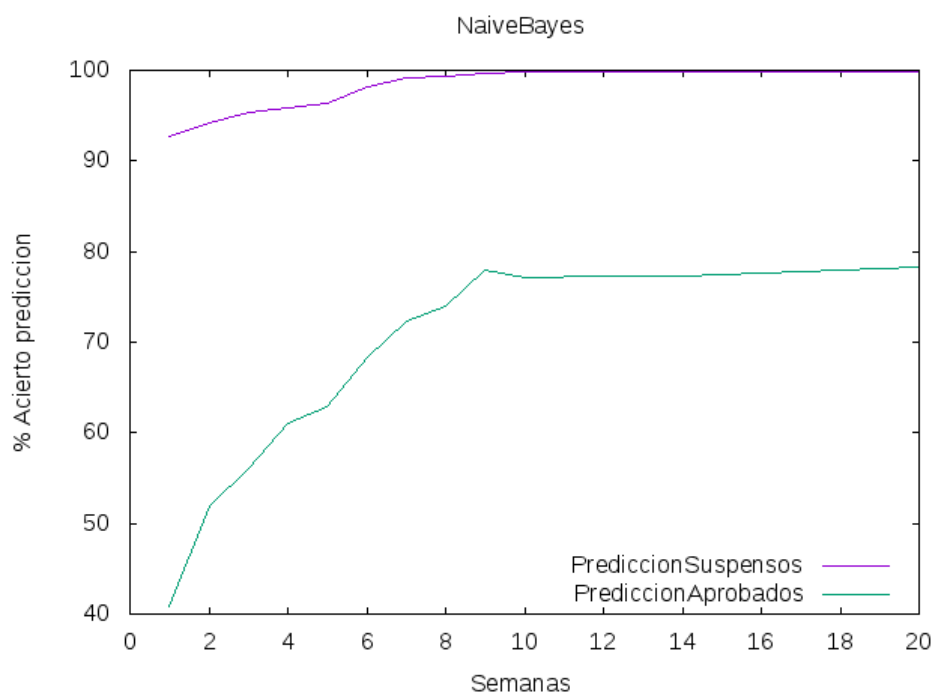


Figura 4.1: Gr  fica Evoluci  n semanal con algoritmo N  ive Bayes en MOOC Don Quijote

4.2.1.2.2. Android

En este apartado se va a exponer la evolución semanal de aplicar el algoritmo de N  ive Bayes en el MOOC Jugando con Android – Aprende a programar tu primera App. En la figura 4.2 se muestra que continua con el mismo patr  n que en el MOOC La Espa  a de Don Quijote. El punto de corte de mejora de predicci  n se encuentra en la semana 8, y en ning  n momento obtenemos una predicci  n aceptable: se acerca a un 70% en la predicci  n de usuarios aprobados, como ocurr  a con los datos totales. La predicci  n de usuarios que suspenden es realmente buena desde la semana 1.

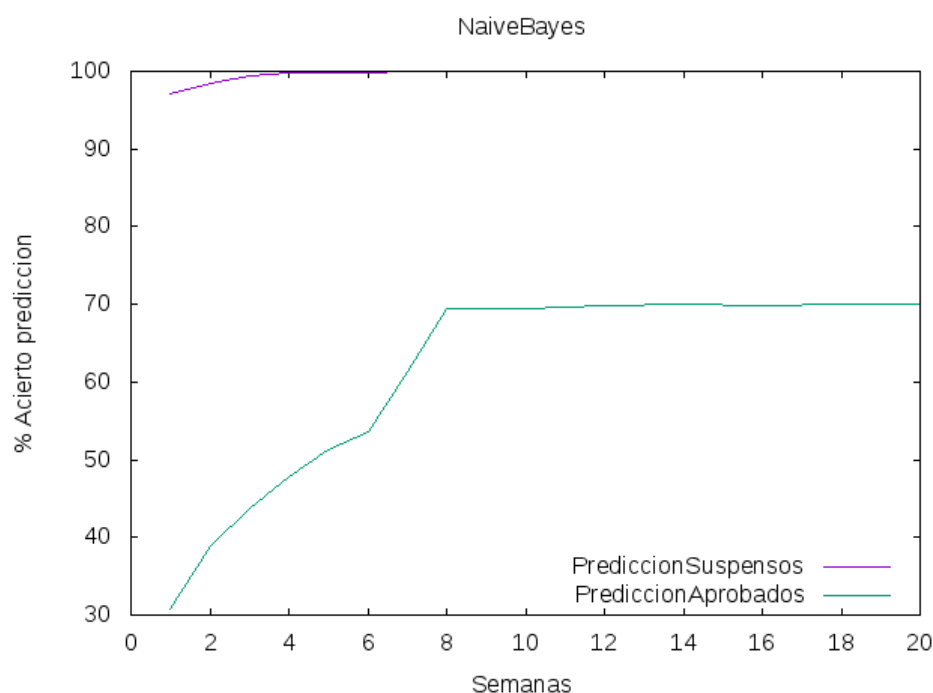


Figura 4.2: Gr  fica Evoluci  n semanal con algoritmo N  ive Bayes en MOOC Android

4.2.2. AdaBoost

El algoritmo AdaBoost (*adaptive boosting*) es un algoritmo de *boosting*. El *boosting* consiste en construir un clasificador muy preciso a partir de m  ltiples clasificadores d  biles [2]. En nuestro caso, cada uno de los atributos que tenemos (n  mero de v  deos, aciertos, intentos etc.) corresponden a un clasificador d  bil. La agregaci  n de cada uno de ellos con su conjunto de entrenamiento forma parte de nuestro algoritmo de AdaBoost.

4.2.2.1. Total

4.2.2.1.1. Don Quijote

En esta secci  n se muestran los resultados de aplicar el algoritmo AdaBoost al total de los datos obtenidos del MOOC La Espa  a de Don Quijote, clasificando los estudiantes en aprobados y suspendidos. En la siguiente tabla se pueden observar los resultados.

<i>Alumnos</i>	<i>Suspense</i>	<i>Aprobado</i>
<i>Suspense</i>	1748	16
<i>Aprobado</i>	3	196
<i>Porcentaje</i>	<i>Suspense</i>	<i>Aprobado</i>
<i>Suspense</i>	99.8	7.5
<i>Aprobado</i>	0.2	92.5

Tabla 4.3: Tabla de alumnos y porcentaje de alumnos aprobados y suspensos en MOOC de Don Quijote con algoritmo AdaBoost

En la tabla 4.3 podemos observar que clasifica realmente bien a aquellos alumnos para los que la predicción es que suspenden. Pero el dato importante es la mejora que produce con respecto al algoritmo N  ive Bayes en la predicci  n de los alumnos que deber  an aprobar. Aun as  , esta mejora no es del todo satisfactoria, ya que un 7.5% de los alumnos clasificados en la predicci  n como aprobados, suspenden.

4.2.2.1.2. Android

En este apartado se muestran los resultados de aplicar el algoritmo AdaBoost total de los datos obtenidos del MOOC Jugando con Android – Aprende a programar tu primera App, clasificando los estudiantes en aprobados y suspendidos. En la siguiente tabla se pueden observar los resultados.

<i>Alumnos</i>	<i>Suspense</i>	<i>Aprobado</i>
<i>Suspense</i>	7999	44
<i>Aprobado</i>	9	484
<i>Porcentaje</i>	<i>Suspense</i>	<i>Aprobado</i>
<i>Suspense</i>	99.9	8.3
<i>Aprobado</i>	0.1	91.7

Tabla 4.4: Tabla de alumnos y porcentaje de alumnos aprobados y suspensos en MOOC de Android con algoritmo AdaBoost

La tabla 4.4 confirma la conclusi  n observada en la tabla 4.3: la predicci  n para los alumnos que suspenden es realmente buena, pero pese a la mejora con respecto al algoritmo N  ive Bayes, el error cometido es de un 8.3% de los alumnos que se clasifican como aprobados en la predicci  n.

4.2.2.2. Semanal

4.2.2.2.1. Don Quijote

En este apartado se va a exponer la evoluci  n semanal de aplicar el algoritmo AdaBoost en el MOOC La Espa  a de Don Quijote. En la figura 4.3 se muestra una gr  fica que choca con lo que hasta ahora hemos visto. Las gr  ficas del algoritmo N  ive Bayes manten  an una subida constante hasta llegar a una semana en la que pr  cticamente alcanzaba su m  ximo. En este caso, observamos que en las primeras dos semanas obtenemos un acierto al predecir los usuarios aprobados cercana al 78%, pero este porcentaje cae por debajo del 70% en las semanas 3, 4 y 5.

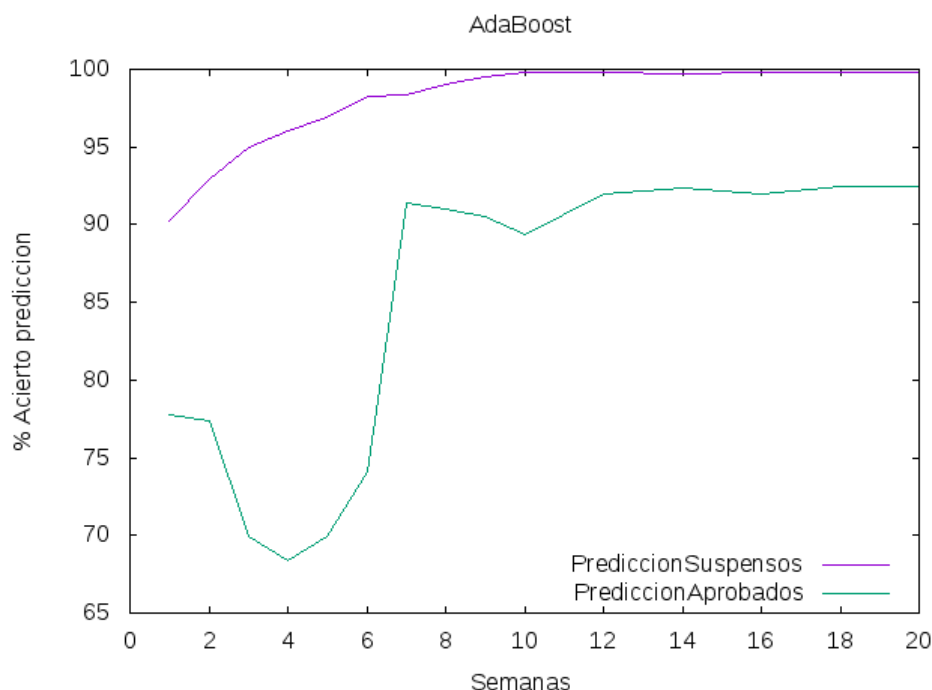


Figura 4.3: Gráfica Evolución semanal con algoritmo AdaBoost en MOOC Don Quijote

La explicación a esto es que la tasa de acierto de suspensos mejora notablemente, con lo que podemos afirmar que muchos de los que finalmente aprobaran estas semanas no han empezado con la fuerza suficiente como para diferenciarlos de los usuarios sin interés. Por ello hay un número muy reducido de catalogados como aprobados y este número tan reducido proporciona una buena predicción para ser las dos primeras semanas. Sin embargo, a partir de la semana 3, aumenta el número de alumnos con predicción de aprobados, por lo que los fallos en la predicción de suspensos mejoran, pero provoca que la predicción de alumnos aprobados sea peor.

Alumnos	Semana 1	
	Suspense	Aprobado
Suspense	1762	2
Aprobado	192	7

Tabla 4.5: Tabla de alumnos semana 1 con algoritmo AdaBoost en MOOC Don Quijote

Alumnos	Semana 3	
	Suspense	Aprobado
Suspense	1717	47
Aprobado	107	92

Tabla 4.6: Tabla de alumnos semana 3 con algoritmo AdaBoost en MOOC Don Quijote

Como podemos observar, los datos contenidos en las tablas 4.5 y 4.6 corroboran la explicación dada: el algoritmo es más optimista y por ello comete más errores.

En la semana 7 se produce una subida hasta el 91% en el porcentaje de acierto en la predicción de los usuarios aprobados, llegando al acierto obtenido con el total de los datos. Posteriormente se mantiene sin cambios importantes.

4.2.2.2.2. Android

En este apartado se va a exponer la evolución semanal de aplicar el algoritmo AdaBoost en el MOOC Jugando con Android – Aprende a programar tu primera App. En la figura 4.4 se observa algo parecido a lo que ocurría en la figura 4.3 del apartado anterior, pero en este caso la bajada de porcentaje de acierto ocurre en la semana 4. Se aplica la misma explicación, como podemos ver en las tablas 4.7 y 4.8 correspondientes a la semana 3 y 4 respectivamente. Además, vemos que en la semana 1 hay un acierto del 0% debido a que no clasifica a ningún alumno como aprobado.

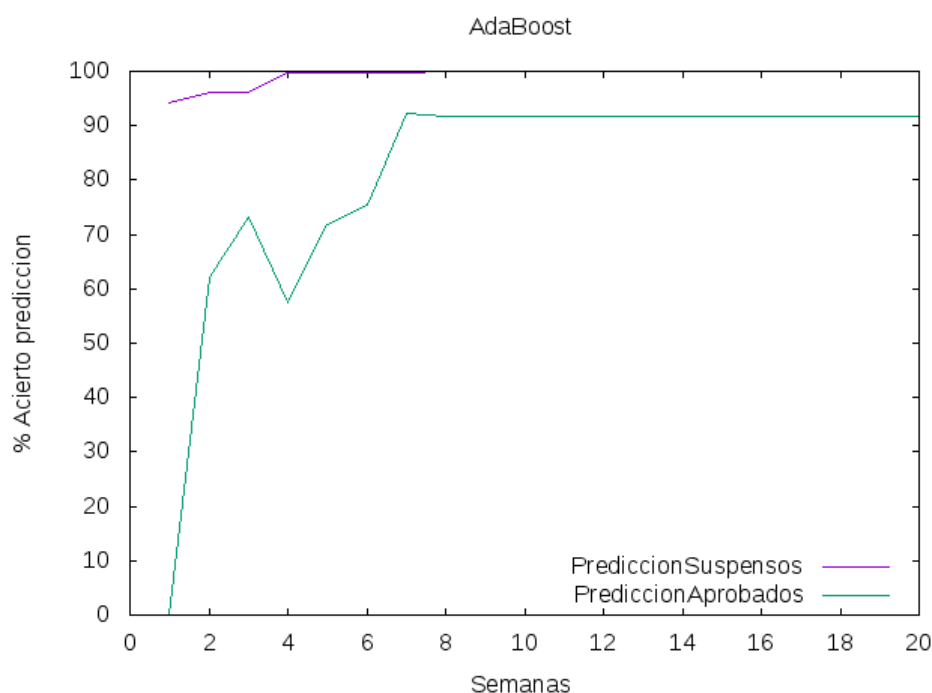


Figura 4.4: Gráfica Evolución semanal con algoritmo AdaBoost en MOOC Android

En la semana 7 obtenemos el punto en el que se estanca la subida de porcentaje de acierto, llegando prácticamente al 92% conseguido con todos los datos.

	Semana 3	
Alumnos	Suspense	Aprobado
Suspense	7983	60
Aprobado	328	165

Tabla 4.7: Tabla de alumnos semana 3 con algoritmo AdaBoost en MOOC Android

	Semana 4	
Alumnos	Suspense	Aprobado
Suspense	7695	348
Aprobado	23	470

Tabla 4.8: Tabla de alumnos semana 4 con algoritmo AdaBoost en MOOC Android

Evaluando estos datos vemos que la tendencia es la misma que con los datos totales. Se puede asegurar que semanalmente también es mejor AdaBoost que N  ive Bayes para la predicci  n de los resultados de los alumnos.

4.2.3. ADTree

El algoritmo ADTree (*Alternating Decision Tree*) es un   rbol de decisi  n alternativo. Este tipo de   rboles contiene dos tipos de nodos: los nodos de predicci  n y los nodos de decisi  n. Los nodos de predicci  n contienen solo un n  mero o valor. Los de decisi  n contienen un predicado sobre el cual se eval  an las instancias. Seg  n atraviesas los nodos de decisi  n, se sumará el valor del nodo de precisi  n que se atraviesa. Para entender mejor este concepto se mostrar   en un ejemplo.

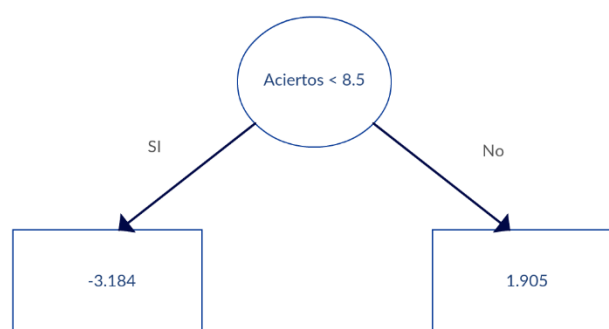


Figura 4.5: Representaci  n de los nodos de decisi  n y predicci  n del algoritmo ADTree

En la figura 4.5 se representa el nodo de decisi  n con una elipse y los de predicci  n con un rect  ngulo. En funci  n del camino cogido tras la pregunta en el nodo de decisi  n, se llega a un nodo de predicci  n que contiene un valor que se sumará. Se ha representado un simple nodo, pero puede haber muchos nodos que se sumen entre ellos para obtener un valor final. Este valor final decidir   si la instancia se catalogar   como alumno suspenso o alumno aprobado.

4.2.3.1. Total

4.2.3.1.1. Don Quijote

En esta secci  n se muestran los resultados de aplicar el algoritmo ADTree al total de los datos obtenidos del MOOC La Espa  a de Don Quijote, clasificando los estudiantes en aprobados y suspendidos. En la siguiente tabla se pueden observar los resultados.

Alumnos	Suspenso	Aprobado
Suspenso	1756	8
Aprobado	3	196
Porcentaje	Suspenso	Aprobado
Suspenso	99.8	3.9
Aprobado	0.2	96.1

Tabla 4.9: Tabla de alumnos y porcentaje de alumnos aprobados y suspensos en MOOC de Don Quijote con algoritmo ADTree

En la tabla 4.9 comprobamos, una vez m  s, que la predicci  n de los alumnos que van a suspender no es un problema. El algoritmo ADTree acierta un 96.1% de los alumnos que

van a aprobar, es decir, solo comete un error de un 3.9% para este MOOC. Se puede decir por ello que mejora al algoritmo AdaBoost, siendo hasta ahora el mejor clasificador.

4.2.3.1.2. Android

En este apartado se muestran los resultados de aplicar el algoritmo ADTree total de los datos obtenidos del MOOC Jugando con Android – Aprende a programar tu primera App, clasificando los estudiantes en aprobados y suspendidos. En la siguiente tabla se pueden observar los resultados.

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	8026	17
<i>Aprobado</i>	12	481
<i>Porcentaje</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	99.9	3.4
<i>Aprobado</i>	0.1	96.6

Tabla 4.10: Tabla de alumnos y porcentaje de alumnos aprobados y suspensos en MOOC de Android con algoritmo ADTree

La tabla 4.10 nos confirma los resultados de la tabla 4.9. Se continua con un error despreciable en las predicciones de alumnos que suspenden y comete un error de un 3.4% en la predicción de alumnos que aprueban, bajando incluso el porcentaje de fallo para este MOOC.

4.2.3.2. Semanal

4.2.3.2.1. Don Quijote

En este apartado se va a exponer la evolución semanal de aplicar el algoritmo ADTree en el MOOC La España de Don Quijote. En la figura 4.6 podemos ver que la predicción de los suspensos es similar a los datos obtenidos en AdaBoost y N  ive Bayes. La predicci  n de aprobados aumenta tras la semana 4; hasta la semana 3 obtiene un porcentaje inferior de 75% de acierto, pasando a un porcentaje superior al 90 % en la semana 4. En semanas posteriores no mantiene una regularidad tan marcada como con los dos algoritmos anteriores. Aun con esta irregularidad a partir de la semana 4, se puede decir que es mejor que el algoritmo AdaBoost.

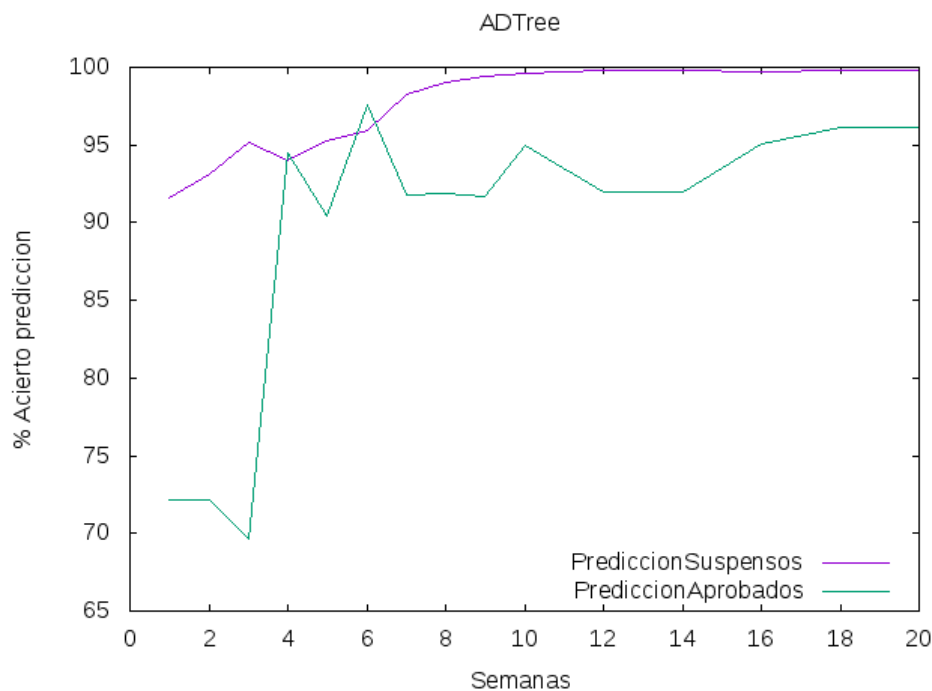


Figura 4.6: Grafica Evolución semanal con algoritmo ADTree en MOOC Don Quijote

4.2.3.2.2. *Android*

En este apartado se va a exponer la evolución semanal de aplicar el algoritmo ADTree total de los datos obtenidos del MOOC Jugando con Android – Aprende a programar tu primera App. En la figura 4.7 observamos que presenta un crecimiento continuado hasta las semanas 7-8, en las que alcanza el porcentaje máximo, logrando hasta un 95%. Obtiene un mejor rendimiento que AdaBoost.

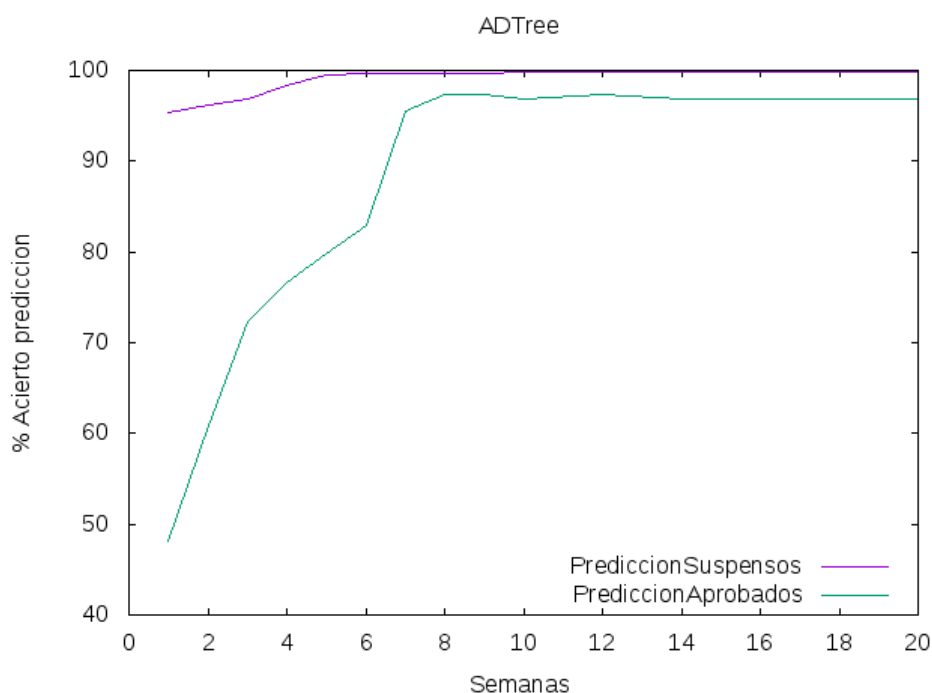


Figura 4.7: Grafica Evolución semanal con algoritmo ADTree en MOOC Android

En este caso, no se refleja la variación que tiene la figura 4.6, por lo que vemos que actúa de manera diferente en los dos MOOCs estudiados.

4.2.4. J48

J48 es una implementación en código Java del algoritmo C4.5. C4.5 crea árboles de decisión a partir de un grupo de datos de entrenamiento usando el concepto de entropía de información. En los nodos del árbol se elige el atributo que mayor ganancia de información proporciona para separar en subconjuntos los datos, dando lugar a estos subconjuntos más pequeños que pasaran a otro nodo para continuar la operación. El algoritmo termina cuando el árbol llega a un punto donde todas las instancias coinciden en una clase.

4.2.4.1. Total

4.2.4.1.1. Don Quijote

En esta sección se muestran los resultados de aplicar el algoritmo J48 al total de los datos obtenidos del MOOC La España de Don Quijote, clasificando los estudiantes en aprobados y suspendidos. En la siguiente tabla se pueden observar los resultados.

<i>Alumnos</i>	<i>Suspense</i>	<i>Aprobado</i>
<i>Suspense</i>	1758	6
<i>Aprobado</i>	4	195
<i>Porcentaje</i>	<i>Suspense</i>	<i>Aprobado</i>
<i>Suspense</i>	99.8	3
<i>Aprobado</i>	0.2	97

Tabla 4.11: Tabla de alumnos y porcentaje de alumnos aprobados y suspensos en MOOC de Don Quijote con algoritmo J48

En la tabla 4.11 se puede observar que la predicción de alumnos suspensos es muy buena, con un 0.2% de error, y una pequeña mejora de la predicción de alumnos aprobados respecto al algoritmo AdaBoost. Con un 3% de error se puede afirmar que es un buen algoritmo de clasificación y el mejor de los utilizados, a la espera de confirmación en el MOOC Jugando con Android – Aprende a programar tu primera App.

4.2.4.1.2. Android

En este apartado se muestran los resultados de aplicar el algoritmo J48 total de los datos obtenidos del MOOC Jugando con Android – Aprende a programar tu primera App, clasificando los estudiantes en aprobados y suspendidos. En la siguiente tabla se pueden observar los resultados.

<i>Alumnos</i>	<i>Suspense</i>	<i>Aprobado</i>
<i>Suspense</i>	8040	3
<i>Aprobado</i>	11	482
<i>Porcentaje</i>	<i>Suspense</i>	<i>Aprobado</i>
<i>Suspense</i>	99.9	0.6
<i>Aprobado</i>	0.1	99.4

Tabla 4.12: Tabla de alumnos y porcentaje de alumnos aprobados y suspensos en MOOC de Android con algoritmo J48

En la tabla 4.12 se puede observar una mejora para este MOOC en la categoría aprobados, con solo un error de un 0.6% o, lo que es lo mismo, el fallo en la predicción de alumnos aprobados de solo 3 alumnos de 485. No solo confirma la mejora respecto AdaBoost, sino que parece indicar que en MOOCs más numerosos y complejos se puede conseguir una mejor clasificación. Sin duda alguna, se puede afirmar que el algoritmo J48 es el más preciso de los algoritmos de clasificación utilizados en este trabajo.

4.2.4.2. Semanal

4.2.4.2.1. Don Quijote

En este apartado se va a exponer la evolución semanal de aplicar el algoritmo J48 en el MOOC La España de Don Quijote. En la figura 4.8 se puede observar que el porcentaje de predicción de aprobados es un poco caótico, pero una vez más, como en ocasiones anteriores, esto se puede explicar debido a que aumenta el porcentaje de acierto de la predicción de suspensos.

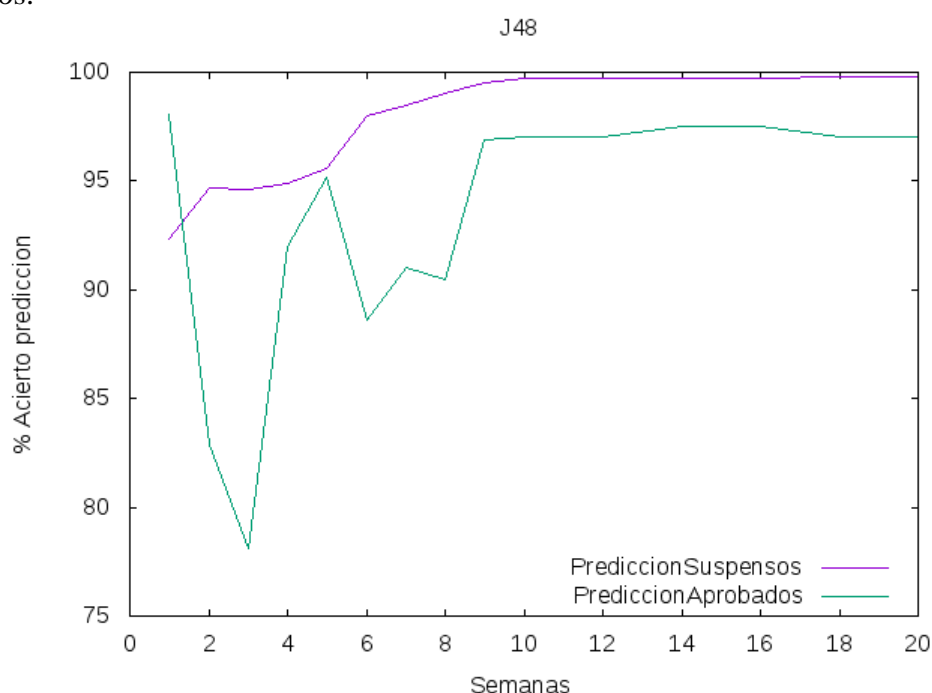


Figura 4.8: Gráfica evolución semanal con algoritmo J48 en MOOC Don Quijote

En las tablas 4.13 y 4.14 podemos observar este suceso: como aumenta el número de alumnos clasificados en aprobados, aumenta el número de errores.

Semana 2		
Alumnos	Suspense	Aprobado
Suspense	1763	1
Aprobado	148	51

Tabla 4.13: Tabla de alumnos semana 2 con algoritmo J48 en MOOC Don Quijote

Alumnos	Semana 3	
	Suspense	Aprobado
Suspense	1743	21
Aprobado	97	102

Tabla 4.14: Tabla de alumnos semana 3 con algoritmo J48 en MOOC Don Quijote

En la semana 9 se alcanza el máximo porcentaje de acierto, con un resultado superior al 95%. J48 obtiene el mejor resultado de los algoritmos utilizados.

4.2.4.2.2. Android

En este apartado se va a exponer la evolución semanal de aplicar el algoritmo J48 total de los datos obtenidos del MOOC Jugando con Android – Aprende a programar tu primera App. En la figura 4.9 podemos observar que con el paso de las semanas va mejorando la predicción de los alumnos aprobados. En la semana 4 se produce un ligero retroceso, una vez más por la causa explicada anteriormente.

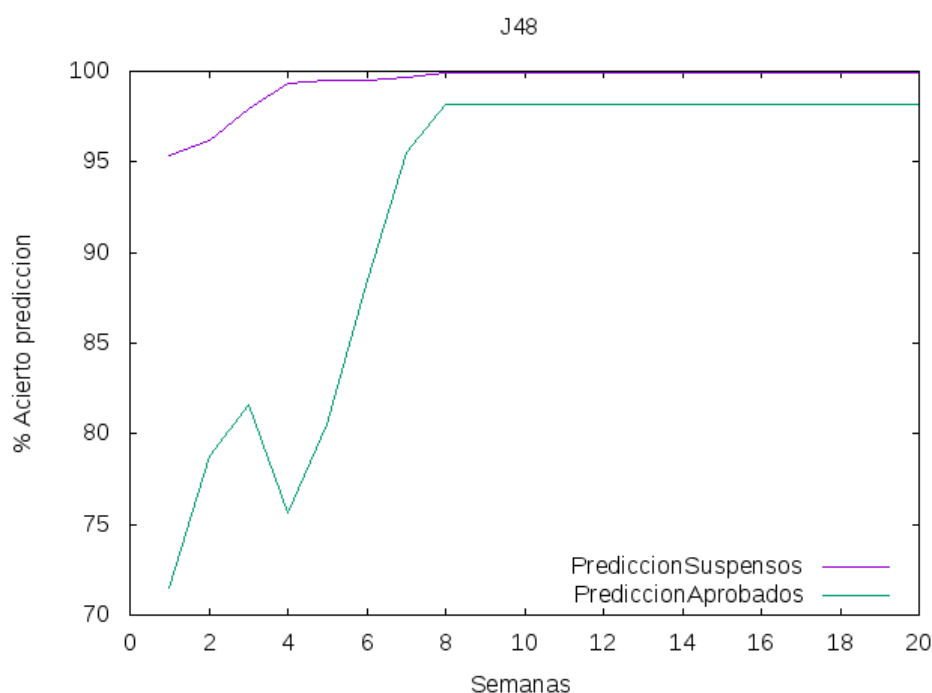


Figura 4.9: Gráfica evolución semanal con algoritmo J48 en MOOC Android

4.2.5. Visión global semanal

En este apartado se comparan los algoritmos utilizados para clasificar los alumnos semanalmente, de modo que se pueda predecir, cada semana, si cada alumno aprobará o suspenderá el curso.

4.2.5.1. Don Quijote

En la figura 4.10 podemos observar que en la predicción de suspensos prácticamente los cuatro algoritmos obtienen unos resultados similares, no se puede deducir cual es mejor o peor. Esto se debe al gran número de acierto, debido al gran porcentaje de abandono de

alumnos a lo largo del curso, que permite detectar fácilmente mediante los diferentes algoritmos a estos alumnos suspensos.

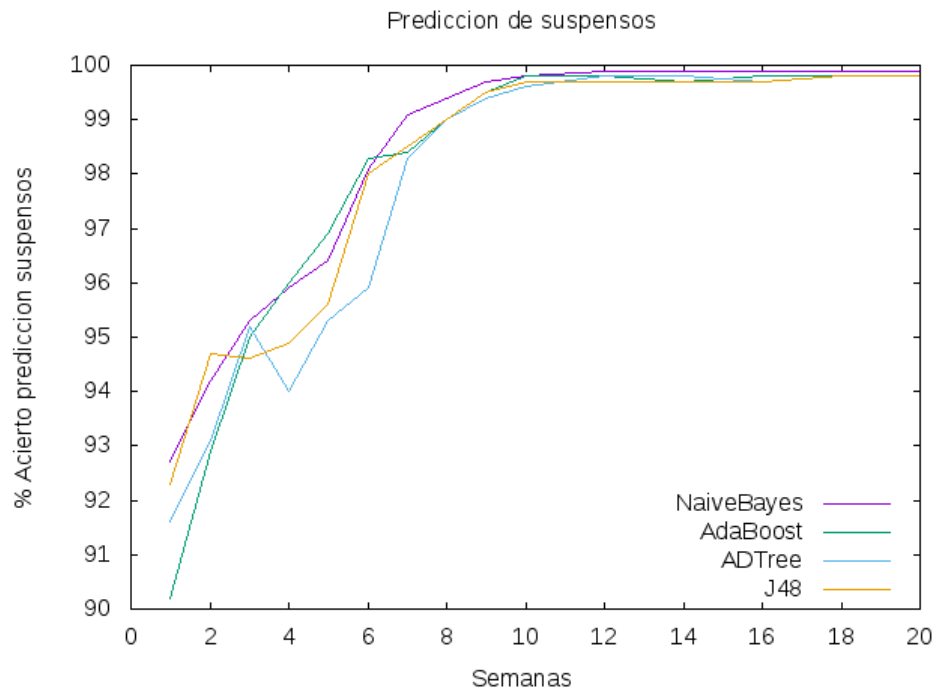


Figura 4.10: Gráfica comparativa predicción de suspensos en MOOC Don Quijote

En la figura 4.11 observamos que el algoritmo N  ive Bayes es el menos eficiente, seguido de AdaBoost. ADTree y J48 tienen un rendimiento parecido en las primeras semanas, a partir de la novena o decima si podemos ver un rendimiento mejor del algoritmo J48.

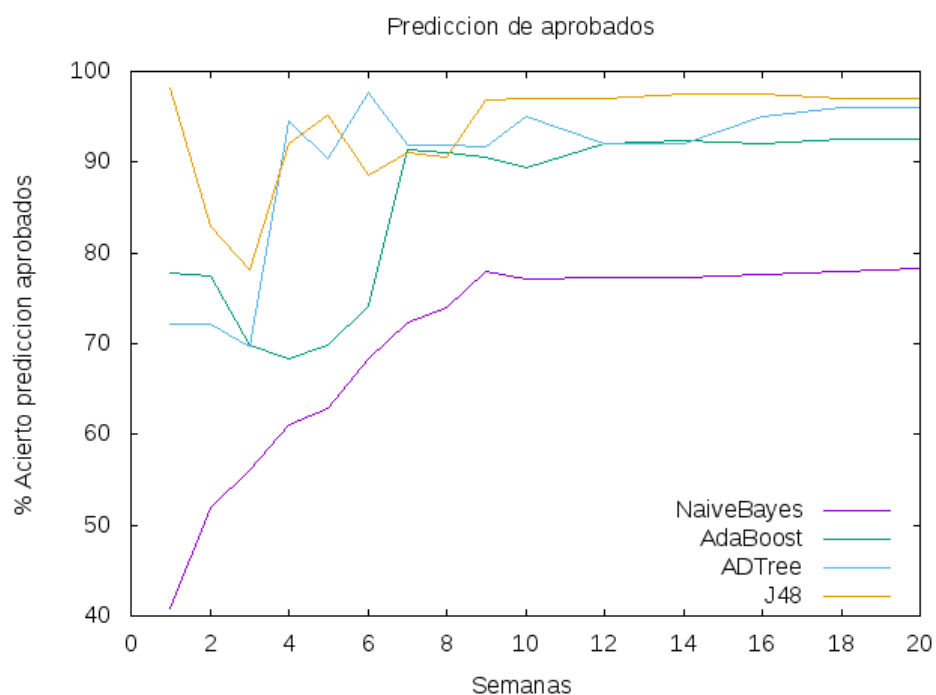


Figura 4.11: Gr  fica comparativa predicci  n de aprobados en MOOC Android

4.2.5.2. Android

En la figura 4.12 podemos observar que en la predicción de suspensos el mejor algoritmo es Naive Bayes, pero esto es debido a la tendencia del algoritmo a clasificar alumnos que están en duda con el aprobado, dejando poco lugar a los fallos en la predicción de suspensos. Entre AdaBoost, ADTree y J48 las diferencias son muy pequeñas, por lo que prácticamente son igual de eficaces.

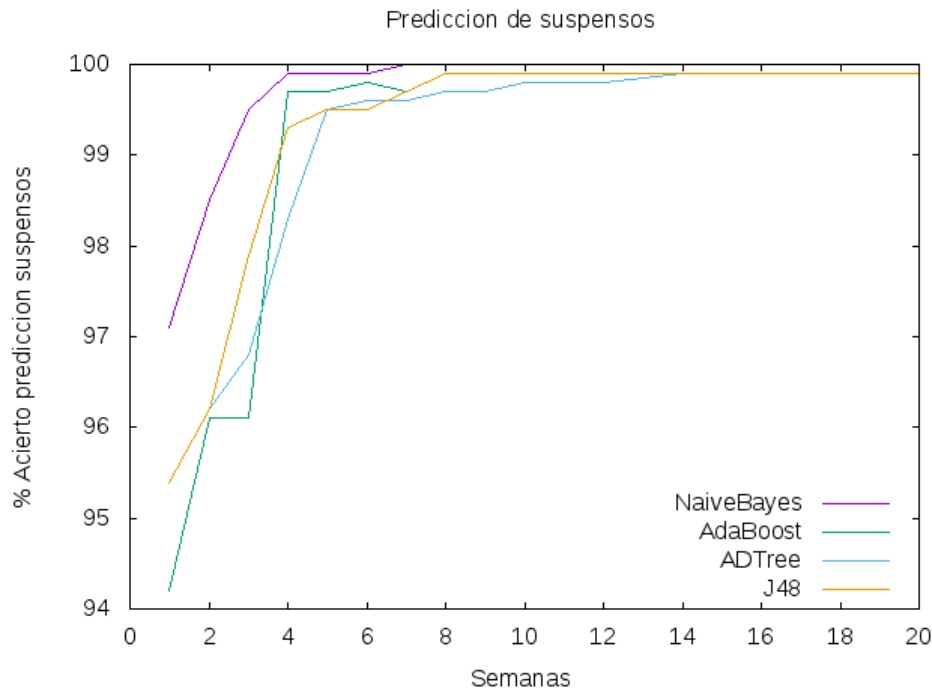


Figura 4.12: Gráfica comparativa predicción de suspensos en MOOC Android

En la figura 4.13, podemos observar como Naive Bayes es el algoritmo que peor predicción hace en la categoría de alumnos aprobados, con una gran diferencia respecto al resto de algoritmos. El algoritmo AdaBoost tiene un rendimiento algo inferior a los algoritmos ADTree y J48 que obtienen un rendimiento similar, aun así, J48 es mínimamente superior en la predicción.

Tras la comparación de los respectivos algoritmos en este apartado y el anterior se deduce que el algoritmo de clasificación con mayor eficiencia es J48.

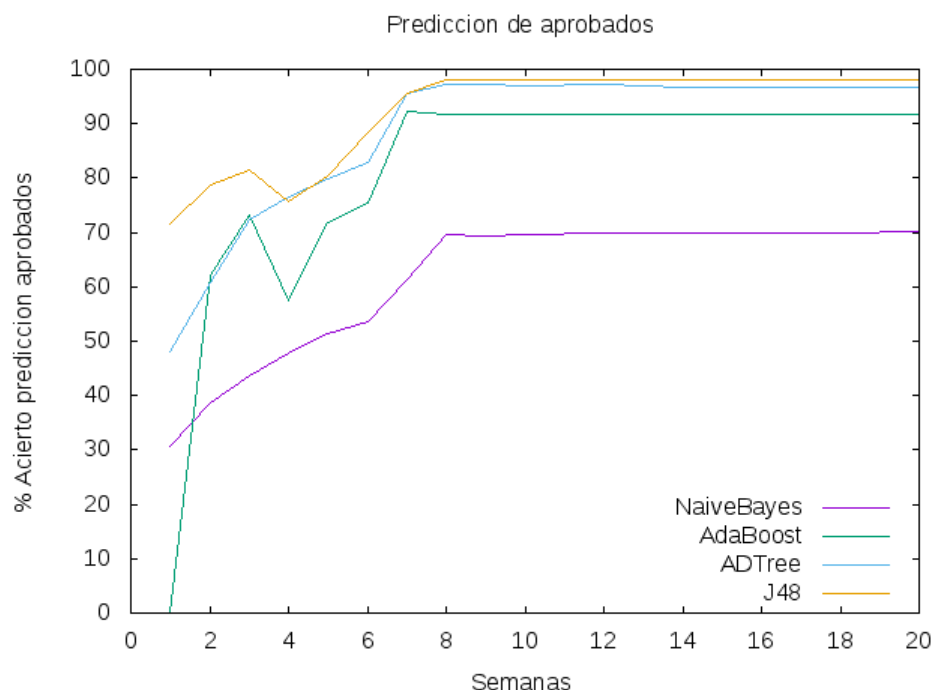


Figura 4.13: Gráfica comparativa predicción de aprobados en MOOC Android

4.3. Clustering

En este apartado se van a analizar los datos y a obtener agrupaciones de los alumnos e información sobre los atributos más relevantes para esas agrupaciones mediante técnicas de *clustering*. Para ello se ha utilizado el algoritmo XMeans, una variante del algoritmo KMeans que da la opción de tener un número de clústeres determinado. Este algoritmo dividirá los alumnos en función de las interacciones que realizan, agrupándoles por similitudes de dichas interacciones, sin determinar a priori el número de clústeres ni identificar algún atributo en particular en el que fijarse para realizar los clústeres.

4.3.1. Don Quijote

4.3.1.1. Total

Como se puede observar en las tablas 4.15 y 4.16, se han obtenido 4 clústeres. El clúster 1 contiene a los alumnos para los que no se alberga ninguna esperanza de éxito. El 100% de los estudiantes incluidos en este clúster suspende. Por tanto, se asume que todos los estudiantes de características similares que en un futuro se clasifiquen en este clúster suspenderán. En el clúster 2 se agrupan aquellos estudiantes para los cuales hay poca esperanza de éxito, pero que no se descarta que aprueben finalmente. Los datos nos muestran que en esta clase hay 3 alumnos de 148 que sí aprueban, cifra que corresponde a un 2% de aprobados dentro del clúster.

Alumnos	Clústeres			
	1	2	3	4
Suspense	1580	145	25	14
Aprobado	0	3	70	126

Tabla 4.15: Tabla de los alumnos de cada clúster MOOC Don Quijote

Porcentaje	Clústeres			
	1	2	3	4
Suspenso	100	98	26.3	10
Aprobado	0	2	73.7	90

Tabla 4.16: Tabla de los alumnos (porcentaje) de cada clúster MOOC Don Quijote

En el clúster 3 se agrupan los alumnos que tienen posibilidades de aprobar, pero que no se puede confirmar con seguridad. En esta tabla se muestra que 70 personas de 95 que pertenecen al clúster 3 aprobaron, lo cual corresponde a un 73.7% de aprobados. Por último, en el clúster 4 se agrupan los alumnos que se estima que van a aprobar. Se puede observar que 126 de 140 aprueban realmente, lo cual corresponde a un 90% de alumnos aprobados dentro del clúster 4. Esta es la clasificación obtenida con el algoritmo de *clustering* utilizado sobre los datos disponibles.

En la figura 4.14 se representa la separación por clústeres. Para esta representación se han elegido los atributos que indican el número de problemas realizados por un alumno (en el eje de abscisas) y el número de eventos registrados para ese alumno (en el eje de ordenadas). El color rojo representa al clúster 1, el azul oscuro al clúster 2, el azul claro al clúster 3 y el verde al clúster 4.

Los alumnos se representan mediante aspas o cuadrados. Un aspa indica que ese estudiante ha sido correctamente incluido en el clúster en el que se le ha clasificado (la mayoría de los alumnos de los clústeres 1 y 4), mientras que un cuadrado significa que para ese estudiante no es correcta la clasificación. En el caso de los clústeres 2 y 3 siempre serán cuadrados debido a que estos clústeres no tienen clase asociada (suspenso o aprobado) al estar en un punto intermedio entre los clústeres 1 y 4.

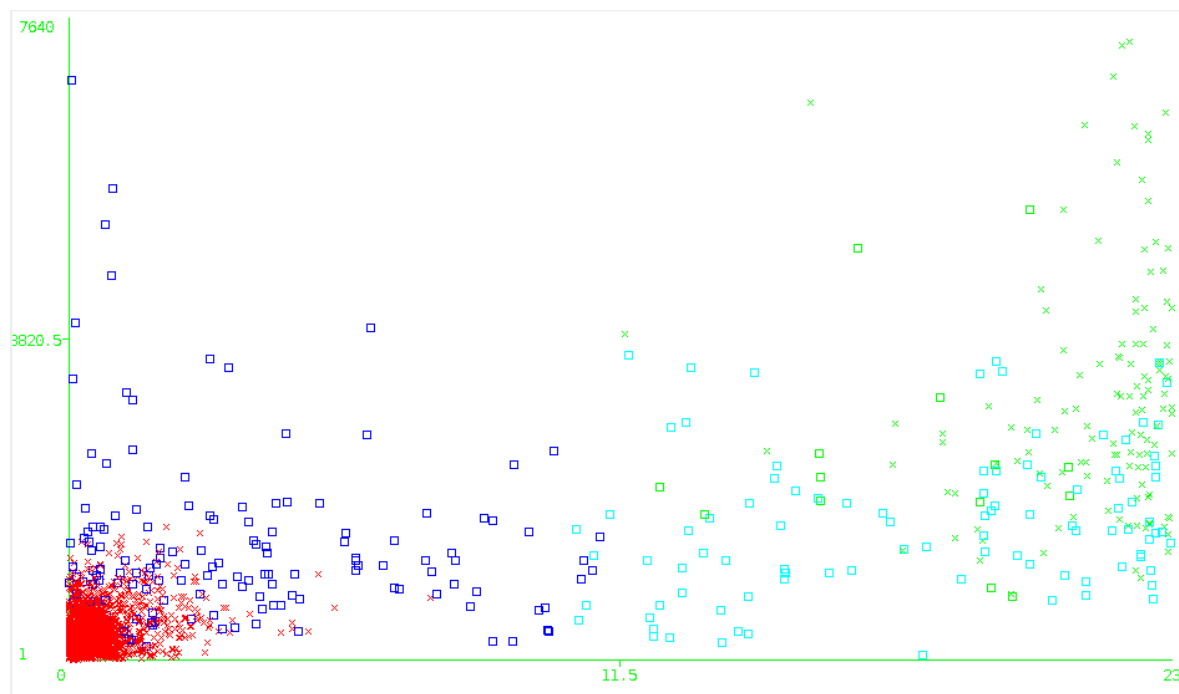


Figura 4.14: Gráfica separación de clústeres problemas-eventos MOOC Don Quijote

Podemos ver cómo están diferenciados claramente los clústeres en función de los eventos y problemas realizados. Cercanos al punto (0,0) se sitúan los estudiantes que han realizado

pocos problemas y pocos eventos en general. Son estudiantes que abandonan el curso pronto. Un poco más a la derecha, en azul oscuro, se sitúan los estudiantes que han dedicado poco esfuerzo pero que no han abandonado en las primeras semanas. La mayoría ha realizado un número escaso de problemas (ninguno llega a 12) y ha generado un número no muy elevado de eventos. Para algunos de ellos sí se ha registrado un número elevado de eventos, pero justamente estos estudiantes han realizado muy pocos problemas; probablemente son estudiantes que están interesados en ver los contenidos del curso, pero no tanto en examinarse/obtener la certificación. En el caso del clúster 4, se puede observar que los estudiantes han realizado casi todos los problemas y han generado un número elevado de eventos. Curiosamente los alumnos suspensos (10%) en este clúster son aquellos que han realizado menos problemas. Y los estudiantes del clúster 3 se sitúan entre el 2 y el 4.

4.3.1.2. Semanal

En la figura 4.15 podemos ver la evolución semanal que tienen los clústeres explicados en el apartado anterior. Es decir, se muestra cómo serían los clústeres cada semana (qué porcentaje de aprobados habría en cada uno de los clústeres en esa semana). Además, se incluye un quinto clúster en las semanas 4 y 5, debido a que en esas semanas los clústeres 2 y 3 se fusionan, dando lugar a un nuevo clúster que engloba a los alumnos que están entremedias entre aprobar y suspender. Se puede observar cómo van evolucionando los clústeres entre las semanas 1 y 6, aproximándose finalmente a los resultados mostrados en el apartado anterior. Se puede observar que desde las primeras semanas el profesor ya tendría datos sobre qué alumnos probablemente aprobarán y cuáles probablemente no lo harán.

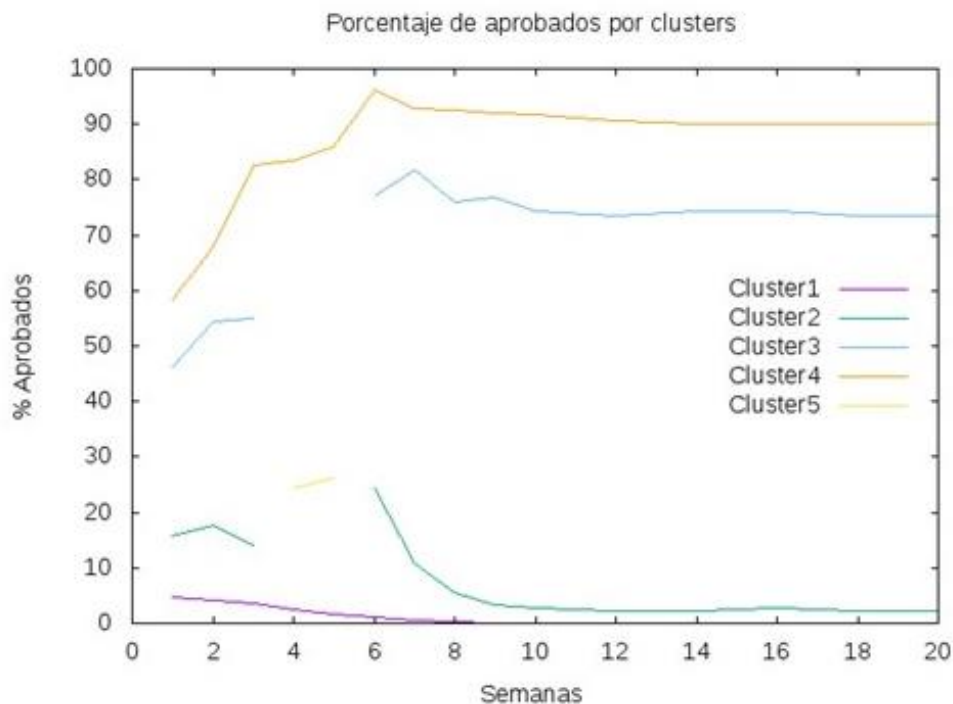


Figura 4.15: Gráfica con porcentaje de alumnos aprobados en cada clúster MOOC Don Quijote

4.3.2. Android

4.3.2.1. Total

Como ocurría en el MOOC La España de Don Quijote, el algoritmo XMeans ha creado 4 clústeres. Los clústeres 1 y 2 representan a aquellos alumnos que prácticamente es imposible que aprueben, aunque hay que hacer una diferencia entre ellos. Los alumnos del clúster 1 tienen una media de 166 eventos en el curso mientras que los del clúster 2 tienen una media de 1169 eventos. La diferencia es muy significativa, y nos indica que pese a tener un resultado final semejante, dado que en ambos las esperanzas de aprobados son nulas (tienen un 100% de alumnos suspendidos), en el clúster 2 al menos son alumnos que se han esforzado algo, mientras que los del clúster 1 se puede decir que simplemente se apuntaron por mera curiosidad, y que no se han esforzado nada.

Alumnos	Clústeres			
	1	2	3	4
<i>Suspense</i>	6155	1363	504	21
<i>Aprobado</i>	0	1	24	468

Tabla 4.17: Tabla de los alumnos de cada clúster MOOC Android

Porcentaje	Clústeres			
	1	2	3	4
<i>Suspense</i>	100	100	95.4	4.3
<i>Aprobado</i>	0	0	4.6	95.7

Tabla 4.18: Tabla de los alumnos (porcentaje) de cada clúster MOOC Android

En el clúster 3 se encuentran los alumnos que están cerca de aprobar pero que no lo han conseguido. En las tablas 4.17 y 4.18 podemos ver que aprueban solamente 24 de 528 alumnos, lo cual corresponde a un 4.6% de aprobados. No es una media muy elevada como para poder diferenciarlo respecto a los clústeres anteriores, pero aun así contiene una diferencia importante en la media de número de eventos realizados por los alumnos; con 2842 supera en más de 1600 la media del clúster 2. Finalmente, en el clúster 4 se agrupan los alumnos que van a aprobar; 468 de 489 lo hacen, esto es un 95.7 de alumnos aprobados dentro de este clúster. En este clúster la media de eventos es de 4688. Se ha tratado de hacer ver la diferencia entre un clúster y otro con el número de eventos, aunque lógicamente no es el único atributo, pero posiblemente sea el que mejor explique las diferencias en cuanto al esfuerzo de los estudiantes de cada clúster.

4.3.2.2. Semanal

En este apartado se pueden ver los clústeres a los que se ha hecho referencia en el apartado anterior por semanas. En la figura 4.16, se puede observar que de la semana 1 a la semana 7 se producen cambios, y a partir de ahí los cambios son muy pequeños. Esto quiere decir que los usuarios tardan unas 7-8 semanas en realizar el curso, por lo menos aquellos que tienen interés. La tendencia de los clústeres 1 y 2 es acercarse a ese 0% de aprobados: desde la semana 1 tienen un 1% y 6% de aprobados respectivamente, aunque hay que decir que “tiene trampa”. Un gran número de usuarios se apuntan y realmente no tienen interés en el curso. Por ello la tasa de acierto se ajusta tanto a lo que los clústeres indican. En esta primera semana estos dos clústeres tienen 66 y 119 alumnos, respectivamente, que finalmente aprobarán. Los clústeres 3 y 4 en la primera semana tienen un 25% y un 43% de acierto

respectivamente. Estos dos clústeres son interesantes, pues representan aquellos alumnos con oportunidades reales de aprobar. Lo curioso es que el clúster 3 con el paso de las semanas tiende a reducir su porcentaje de alumnos aprobados, contrariamente a lo que pasaba en el MOOC La España de Don Quijote. El clúster 4 sí que obtiene un aumento del porcentaje de alumnos aprobados con el paso de las semanas, aumentando el porcentaje de alumnos aprobados prácticamente de manera lineal hasta la semana 7.

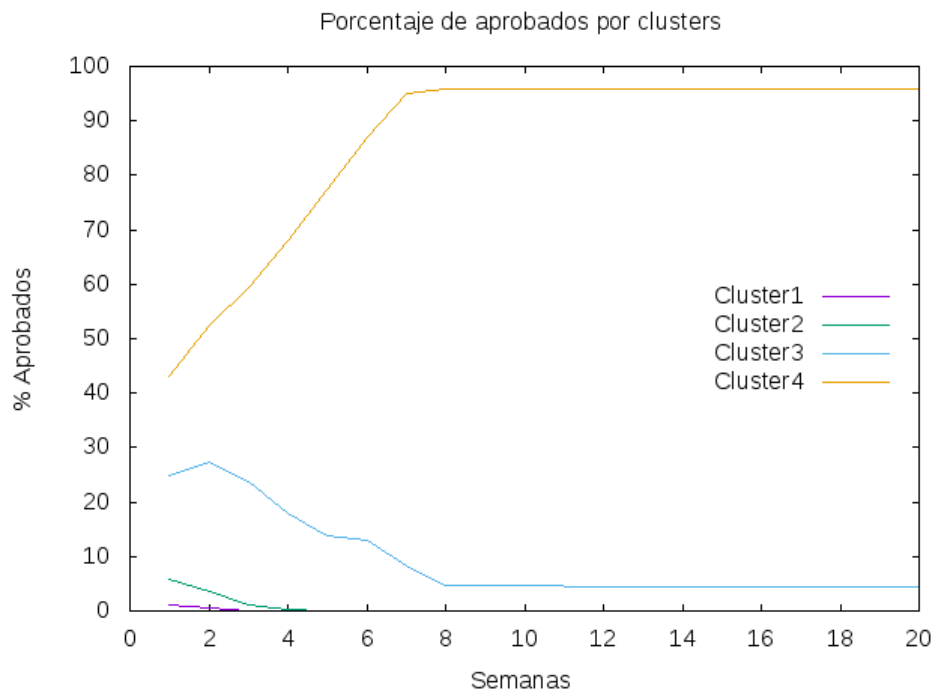


Figura 4.16: Gráfica con porcentaje de alumnos aprobados en cada clúster MOOC Android

4.3.3. Comparación de datos

Se ha podido observar que los dos MOOCs no proporcionan una igualdad en los clústeres. Se puede observar cómo en MOOC de Don Quijote existen dos clústeres con los alumnos que tienden a aprobar y otros dos clústeres con alumnos que tienden a suspender. Sin embargo, en el MOOC de Android solo un clúster tiende a tener alumnos aprobados. Esto se debe a la dificultad de los MOOCs. El MOOC de Don Quijote tiene una dificultad baja, y por ello no parece necesario estar dentro del clúster de alumnos más aplicados para aprobar, mientras que, en el MOOC de Android, de dificultad elevada, parece muy recomendable formar parte del clúster de alumnos más aplicados para aprobar.

4.4. Combinada

En este apartado se combinan técnicas de *clustering* y clasificación que se han visto anteriormente. Se utilizarán el algoritmo XMeans para *clustering* y el algoritmo J48 para clasificación supervisada, que es el que mostró un mejor rendimiento. Para comprobar si al combinar las dos técnicas se mejora el rendimiento vamos a comparar los resultados obtenidos con la combinación con los obtenidos con el algoritmo de clasificación J48. Además, debido a que los resultados tienden a estabilizarse a partir de las semanas 8 o 9, vamos a acotar el estudio hasta la semana 10.

4.4.1. Don Quijote

En la figura 4.17, que representa el porcentaje de acierto en la predicción del suspenso cada semana para ambas aproximaciones, se observa que hay muy poca diferencia entre ellas, y dependiendo la semana puede ser mejor el método combinado o el algoritmo J48 en cuanto al porcentaje de acierto de suspensos. Aun así, consigue un ligero mejor rendimiento el método combinado.

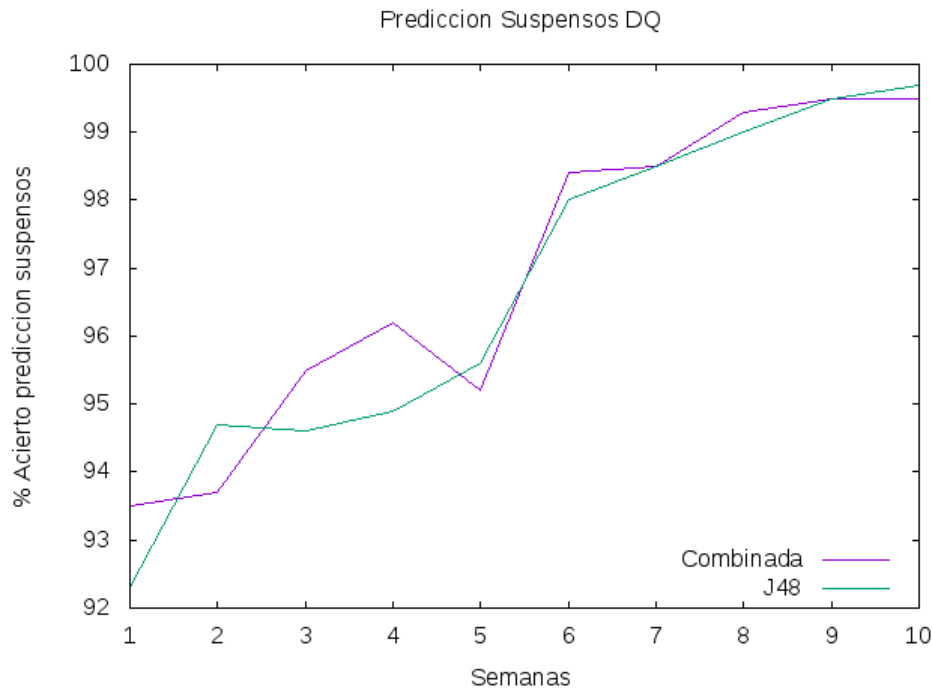


Figura 4.17: Gráfica comparación predicción de suspensos en MOOC Don Quijote

En la figura 4.18, que representa el porcentaje de acierto en la predicción de aprobados, sí se aprecia una mejora del método combinado, con la excepción de las semanas 1 y 5. En la semana 3 el método combinado consigue 6 puntos porcentuales por encima del algoritmo J48, en la semana 4 consigue 2.9 puntos porcentuales. En la semana 5 es mejor J48 en 1.9 puntos porcentuales. A partir de la semana 6 la combinada es mejor.

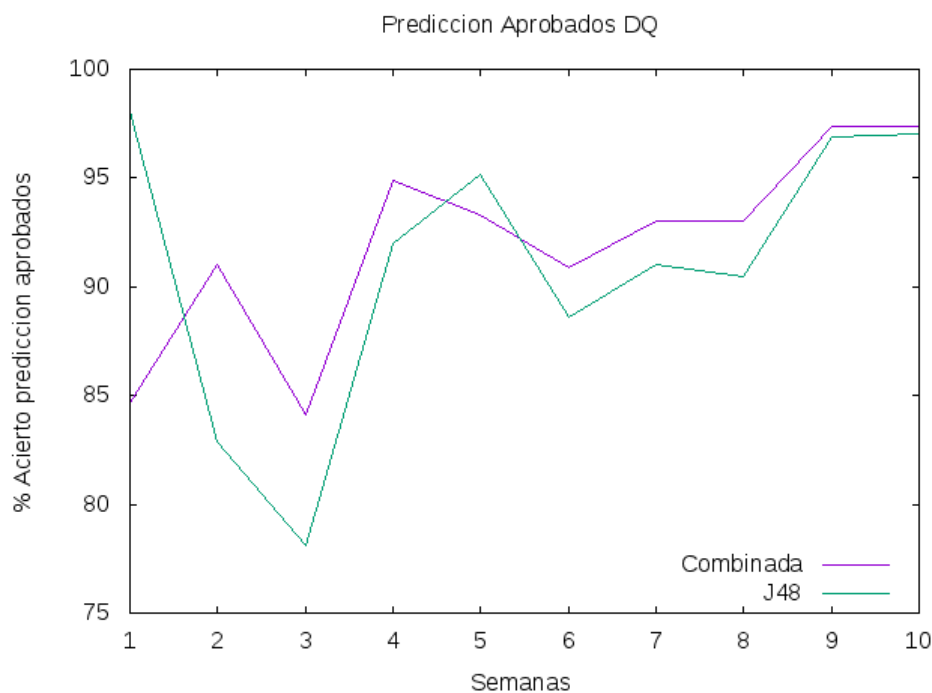


Figura 4.18: Gráfica comparación predicción de aprobados en MOOC Don Quijote

4.4.2. Android

En la figura 4.19 observamos que los resultados de ambos métodos para la predicción de suspensos son bastante similares. Aun así, cabe remarcar que el método combinado obtiene un mejor resultado de la semana 1 a la 3 y de las semanas 5 y 7; por el contrario, en la semana 4 obtiene una mejor predicción el algoritmo J48.

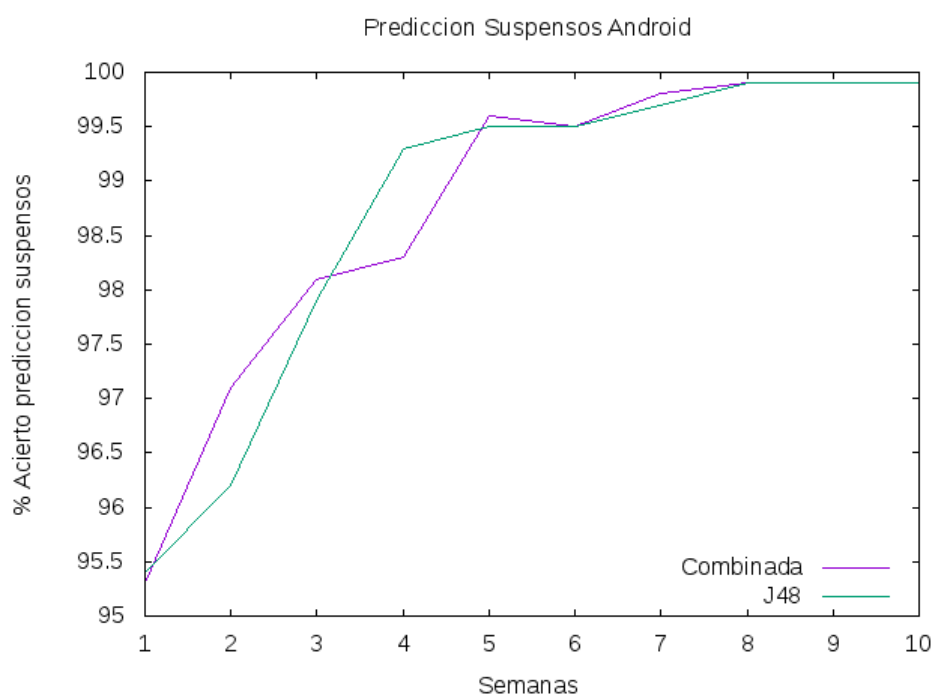


Figura 4.19: Gráfica comparación predicción de suspensos en MOOC Android

En la figura 4.20 vemos que la predicción en las tres primeras semanas es mejor en el caso de J48, pero entre las semanas 4 y 7 es mejor la combinada. En este caso nos aporta más información la combinada debido a que en las semanas más próximas a examinarse tiene un mayor porcentaje de acierto.

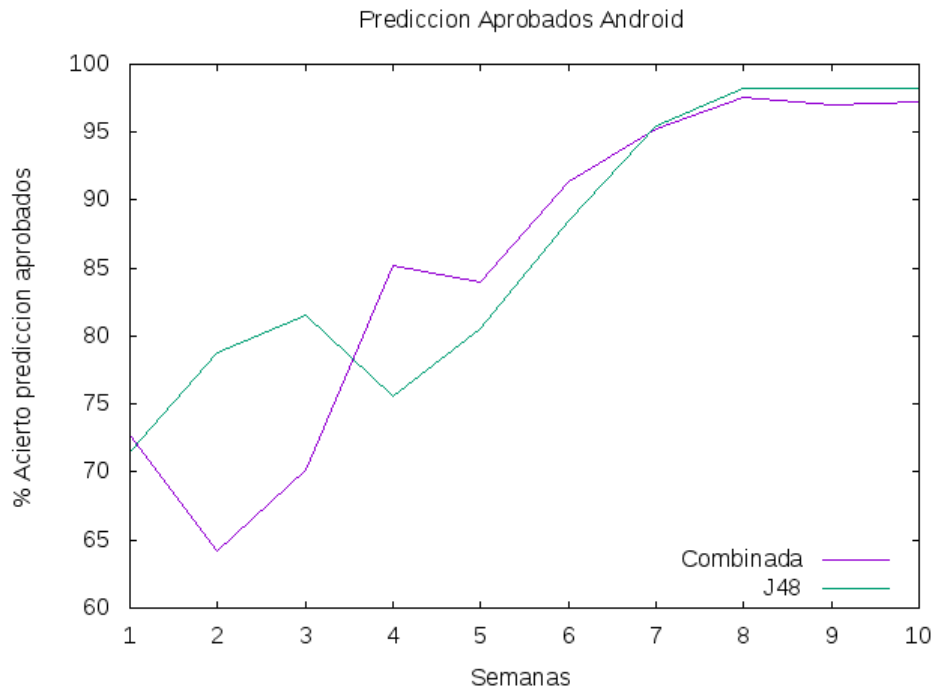


Figura 4.20: Gráfica comparación predicción de aprobados en MOOC Android

Entre las semanas 1 y 3 el método combinado es más optimista de cara a los alumnos, otorgando a un mayor número de ellos la predicción de aprobados. En la semana 4 sucede lo contrario, el algoritmo J48 es más optimista y por ello concede un mayor error a la hora de predecir los alumnos aprobados. A partir de la semana 5 se puede considerar que tienen un optimismo similar ambos métodos, mostrando una igualdad en el porcentaje de éxito, siendo ligeramente mejor el método combinado.

4.4.3. Comparación de datos

Tras la realización del estudio comparativo entre las dos aproximaciones propuestas, se pone de manifiesto una circunstancia que ya se había comentado anteriormente: dependiendo de la dificultad del MOOC que se evalúa con estos métodos, podría cambiar el rendimiento de los mismos. En el MOOC de Don Quijote se puede observar que el método combinado obtiene mejores resultados que el algoritmo de clasificación J48. Sin embargo, en el MOOC de Android prácticamente obtienen el mismo rendimiento. Aplicar *clustering* es más beneficioso para MOOCs de una dificultad reducida.

5 Conclusiones y trabajo futuro

5.1. Conclusiones

En este trabajo se ha llevado a cabo un estudio de diferentes métodos para dar soporte al análisis de la información sobre el uso de los MOOCs por parte de los estudiantes, para tratar de predecir el riesgo de fracaso o abandono y poder informar a los profesores semanalmente para que se tomen las medidas que se consideren oportunas (por ejemplo, motivarles a los estudiantes que van bien o regular, para que continúen trabajando, o indicar a los alumnos que van a suspender que si quieren conseguir el certificado deben trabajar más).

Con algoritmos de aprendizaje automático se ha procedido a la clasificación de los estudiantes a partir de los datos disponibles. Se ha trabajado con algoritmos de clasificación supervisada, no supervisada y una combinación de ambos. Dentro de los algoritmos de clasificación supervisada, el que mejor rendimiento ha obtenido es J48, por ello se recomienda su uso por delante de Naïve Bayes, AdaBoost y ADTree. Con respecto a la clasificación no supervisada, utilizando *clustering* con XMeans se han obtenido 4 y, en ocasiones puntuales, 3 perfiles de alumnos. Estos 4 perfiles van desde el alumno que ha abandonado el curso, pasando por el alumno que tiene interés pero no está realizando el suficiente esfuerzo y por el alumno que tiene interés y se está esforzando, hasta el alumno que está totalmente involucrado en el curso y su nivel de esfuerzo es elevado y por ello no se espera que tenga problemas para conseguir el certificado. Un dato interesante de la utilización de este método es que se obtienen diferentes resultados en función de la dificultad del MOOC. En un MOOC de dificultad elevada parece necesario estar en el perfil de estudiante más aplicado para conseguir el certificado. Por el contrario, en un MOOC de dificultad baja no es necesario que los alumnos estén dentro del grupo de alumnos punteros para aprobar. Con esta información, si el profesor es consciente del grado de dificultad de su MOOC, puede adecuar el nivel de exigencia semanal que pide a sus alumnos.

Finalmente se ha procedido a combinar estos dos métodos. La conclusión es que mejora sensiblemente la predicción en MOOCs cuya dificultad es baja. Sin embargo, en MOOCs con una dificultad elevada los resultados obtenidos son prácticamente iguales a los que ofrece el algoritmo de clasificación J48. En global sería recomendable utilizar este método, ya que mejora en algunos casos y en otros iguala el rendimiento, pero no empeora.

5.2. Trabajo futuro

Como trabajo futuro sería interesante comprobar si las conclusiones extraídas se pueden generalizar, realizando el mismo estudio en otros MOOCs. Además, sería interesante realizar un estudio de *clustering* explorando otros algoritmos. En este trabajo se ha dejado que el algoritmo decidiera el número de clústeres a crear. Podría ser interesante probar a forzar un número distinto de clústeres y comprobar si los resultados mejoran en algún caso.

Sería interesante la creación de un programa que extrajera automáticamente los datos necesarios para realizar este estudio, sin la necesidad de ir al log y filtrar los eventos interesantes para posteriormente conseguir los atributos de cada usuario. En este mismo sentido, se podría aumentar el número de atributos utilizados.

Sería muy interesante también llevar a cabo un estudio mientras los alumnos están realizando el MOOC, en tiempo real. De este modo se podría preguntar a los profesores la utilidad de la información proporcionada semanalmente sobre las predicciones, y estudiar qué se podría

mejorar, el modo en que se les podría ofrecer la información, qué datos necesitarían y si tendría sentido implementar un sistema de alertas automática para los estudiantes y cuáles serían sus requisitos, con la intención de desarrollarlo todo en un futuro.

En esta dirección, se plantea como trabajo futuro desarrollar una aplicación informática con una interfaz adecuada para mostrar la información manejada en este estudio, las predicciones semanales, las alertas sobre el riesgo de fracaso o abandono de los estudiantes, etc., que pudiera ser manejada por los responsables de los MOOCs. También se podría explorar la posibilidad de incorporar alertas para los estudiantes en los propios MOOCs.

Referencias

- [1] “12. Events in the Tracking Logs¶.” 12. Events in the Tracking Logs - EdX Research Guide Documentation, docstrings.readthedocs.io/en/latest/internal_data_formats/tracking_logs.html#student-events.
- [2] Néstor Paz. “Adaboost Con Aplicación a Detección De Caras Mediante El Algoritmo De Viola-Jones.” 1 Feb. 2009, eva.fing.edu.uy/file.php/514/ARCHIVO/2008/TrabajosFinales2008/NestorPaz2008_informe.pdf.
- [3] Duque, Raúl González. Python Para Todos. Mundogeek.net, 2016.
- [4] “Welcome to Python.org.” Python.org, www.python.org/.
- [5] “Gedit.” Apps/Gedit - GNOME Wiki!, wiki.gnome.org/Apps/Gedit.
- [6] “Weka 3: Data Mining Software in Java.” Weka 3 - Data Mining with Open Source Machine Learning Software in Java, www.cs.waikato.ac.nz/ml/weka/.
- [7] “Word 2016 - Microsoft Store España.” Microsoft - Official Home Page, www.microsoft.com/es-es/store/b/word_2016.
- [8] Staff, EasyBib. “The Free Automatic Bibliography Composer.” EasyBib, Chegg, 1 Jan. 2018, www.easybib.com/.
- [9] “Gnuplot Homepage.” Gnuplot Homepage, www.gnuplot.info/.
- [10] “Massive Open Online Course.” Wikipedia, Wikimedia Foundation, 26 Mar. 2018, es.wikipedia.org/wiki/Massive_Open_Online_Course.
- [11] Becspoc1. "Guía De Iniciación a Los Datos Edx".
- [12] “EdX.” EdX, www.edx.org/.
- [13] Miguel Ángel González-Gallego Sosa, "Predicción y análisis de interacciones de usuarios en plataformas de enseñanza online", junio 2016.
- [14] Colin Taylor, Kalyan Veeramachaneni y Una-May O'Reilly, "Likely to stop? Predicting Stopout in Massive Open Online Courses", agosto 2014.
- [15] Cristina Isidro Estradas, "Propuesta de un método basado en Deep Learning para Learning Analytics en MOOCs", septiembre 2017.
- [16] Angel Perez-Lemonche, Gonzalo Martínez-Muñoz y Estrella Pulido-Cañabate , "Analysing Event Transitions to Discover Student Roles and Predict Grades in MOOCs", 2017.
- [17] “Shotwell.” Projects/GnomeKeyring - GNOME Wiki!, wiki.gnome.org/Apps/Shotwell.

Glosario

AdaBoost	<i>Adaptative boosting</i>
ADTree	<i>Alternating Decision Tree</i>
COMA	Curso On-line Masivo y Abierto
J48	Algoritmo árbol de decisión
MOOC	<i>Massive Open On-line Course</i>
XMeans	Método de agrupamiento

Anexos

A N  ive Bayes

A.1 Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1667	97
<i>Aprobado</i>	132	67

Tabla A.1: Tabla de alumnos semana 1 con algoritmo N  ive Bayes en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1676	88
<i>Aprobado</i>	104	95

Tabla A.2: Tabla de alumnos semana 2 con algoritmo N  ive Bayes en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1672	92
<i>Aprobado</i>	82	117

Tabla A.3: Tabla de alumnos semana 3 con algoritmo N  ive Bayes en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1683	81
<i>Aprobado</i>	72	127

Tabla A.4: Tabla de alumnos semana 4 con algoritmo N  ive Bayes en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1683	81
<i>Aprobado</i>	62	137

Tabla A.5: Tabla de alumnos semana 5 con algoritmo N  ive Bayes en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1687	77
<i>Aprobado</i>	32	167

Tabla A.6: Tabla de alumnos semana 6 con algoritmo N  ive Bayes en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1694	70
<i>Aprobado</i>	16	183

Tabla A.7: Tabla de alumnos semana 7 con algoritmo N  ive Bayes en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1698	66
<i>Aprobado</i>	11	188

Tabla A.8: Tabla de alumnos semana 8 con algoritmo N  ive Bayes en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1709	55
<i>Aprobado</i>	5	194

Tabla A.9: Tabla de alumnos semana 9 con algoritmo N  ive Bayes en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1706	58
<i>Aprobado</i>	4	195

Tabla A.10: Tabla de alumnos semana 10 con algoritmo N  ive Bayes en MOOC Don Quijote

A.2 Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	7431	612
<i>Aprobado</i>	222	271

Tabla A.11: Tabla de alumnos semana 1 con algoritmo N  ive Bayes en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	7446	597
<i>Aprobado</i>	114	379

Tabla A.12: Tabla de alumnos semana 2 con algoritmo N  ive Bayes en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	7459	584
<i>Aprobado</i>	40	453

Tabla A.13: Tabla de alumnos semana 3 con algoritmo N  ive Bayes en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	7515	528
<i>Aprobado</i>	10	483

Tabla A.14: Tabla de alumnos semana 4 con algoritmo N  ive Bayes en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	7579	464
<i>Aprobado</i>	4	489

Tabla A.15: Tabla de alumnos semana 5 con algoritmo N  ive Bayes en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	7618	425
<i>Aprobado</i>	4	489

Tabla A.16: Tabla de alumnos semana 6 con algoritmo N  ive Bayes en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	7733	310
<i>Aprobado</i>	1	492

Tabla A.17: Tabla de alumnos semana 7 con algoritmo N  ive Bayes en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	7828	215
<i>Aprobado</i>	2	491

Tabla A.18: Tabla de alumnos semana 8 con algoritmo N  ive Bayes en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	7826	217
<i>Aprobado</i>	2	491

Tabla A.19: Tabla de alumnos semana 9 con algoritmo N  ive Bayes en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	7828	215
<i>Aprobado</i>	2	491

Tabla A.20: Tabla de alumnos semana 10 con algoritmo N  ive Bayes en MOOC Android

B AdaBoost

B.1 Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1762	2
<i>Aprobado</i>	192	7

Tabla B.21: Tabla de alumnos semana 1 con algoritmo AdaBoost en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1745	19
<i>Aprobado</i>	134	65

Tabla B.22: Tabla de alumnos semana 2 con algoritmo AdaBoost en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1717	47
<i>Aprobado</i>	90	109

Tabla B.23: Tabla de alumnos semana 3 con algoritmo AdaBoost en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1705	59
<i>Aprobado</i>	71	128

Tabla B.24: Tabla de alumnos semana 4 con algoritmo AdaBoost en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1702	62
<i>Aprobado</i>	55	144

Tabla B.25: Tabla de alumnos semana 5 con algoritmo AdaBoost en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1705	59
<i>Aprobado</i>	30	169

Tabla B.26: Tabla de alumnos semana 6 con algoritmo AdaBoost en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1748	16
<i>Aprobado</i>	29	170

Tabla B.27: Tabla de alumnos semana 7 con algoritmo AdaBoost en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1746	18
<i>Aprobado</i>	18	181

Tabla B.28: Tabla de alumnos semana 8 con algoritmo AdaBoost en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1744	20
<i>Aprobado</i>	8	191

Tabla B.29: Tabla de alumnos semana 9 con algoritmo AdaBoost en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1741	23
<i>Aprobado</i>	4	195

Tabla B.30: Tabla de alumnos semana 10 con algoritmo AdaBoost en MOOC Don Quijote

B.2 Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	8043	0
<i>Aprobado</i>	493	0

Tabla B.31: Tabla de alumnos semana 1 con algoritmo AdaBoost en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	7940	103
<i>Aprobado</i>	324	169

Tabla B.32: Tabla de alumnos semana 2 con algoritmo AdaBoost en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	7983	60
<i>Aprobado</i>	328	165

Tabla B.33: Tabla de alumnos semana 3 con algoritmo AdaBoost en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	7695	348
<i>Aprobado</i>	23	470

Tabla B.34: Tabla de alumnos semana 4 con algoritmo AdaBoost en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	7859	184
<i>Aprobado</i>	26	467

Tabla B.35: Tabla de alumnos semana 5 con algoritmo AdaBoost en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	7887	156
<i>Aprobado</i>	18	475

Tabla B.36: Tabla de alumnos semana 6 con algoritmo AdaBoost en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	8004	39
<i>Aprobado</i>	24	469

Tabla B.37: Tabla de alumnos semana 7 con algoritmo AdaBoost en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	7999	44
<i>Aprobado</i>	10	483

Tabla B.38: Tabla de alumnos semana 8 con algoritmo AdaBoost en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	7999	44
<i>Aprobado</i>	9	484

Tabla B.39: Tabla de alumnos semana 9 con algoritmo AdaBoost en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	7999	44
<i>Aprobado</i>	9	484

Tabla B.40: Tabla de alumnos semana 10 con algoritmo AdaBoost en MOOC Android

C ADTree

C.1 Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1749	15
<i>Aprobado</i>	160	39

Tabla C.41: Tabla de alumnos semana 1 con algoritmo ADTree en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1737	27
<i>Aprobado</i>	129	70

Tabla C.42: Tabla de alumnos semana 2 con algoritmo ADTree en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1715	49
<i>Aprobado</i>	87	112

Tabla C.43: Tabla de alumnos semana 3 con algoritmo ADTree en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1759	5
<i>Aprobado</i>	113	86

Tabla C.44: Tabla de alumnos semana 4 con algoritmo ADTree en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1752	12
<i>Aprobado</i>	86	113

Tabla C.45: Tabla de alumnos semana 5 con algoritmo ADTree en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1761	3
<i>Aprobado</i>	75	124

Tabla C.46: Tabla de alumnos semana 6 con algoritmo ADTree en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1749	15
<i>Aprobado</i>	31	168

Tabla C.47: Tabla de alumnos semana 7 con algoritmo ADTree en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1748	16
<i>Aprobado</i>	17	182

Tabla C.48: Tabla de alumnos semana 8 con algoritmo ADTree en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1747	17
<i>Aprobado</i>	10	189

Tabla C.49: Tabla de alumnos semana 9 con algoritmo ADTree en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1754	10
<i>Aprobado</i>	7	192

Tabla C.50: Tabla de alumnos semana 10 con algoritmo ADTree en MOOC Don Quijote

C.2 Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	7926	117
<i>Aprobado</i>	384	109

Tabla C.51: Tabla de alumnos semana 1 con algoritmo ADTree en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	7928	115
<i>Aprobado</i>	314	179

Tabla C.52: Tabla de alumnos semana 2 con algoritmo ADTree en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	7957	86
<i>Aprobado</i>	267	226

Tabla C.53: Tabla de alumnos semana 3 con algoritmo ADTree en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	7934	109
<i>Aprobado</i>	136	357

Tabla C.54: Tabla de alumnos semana 4 con algoritmo ADTree en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	7928	115
<i>Aprobado</i>	41	452

Tabla C.55: Tabla de alumnos semana 5 con algoritmo ADTree en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	7947	96
<i>Aprobado</i>	29	464

Tabla C.56: Tabla de alumnos semana 6 con algoritmo ADTree en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	8021	22
<i>Aprobado</i>	30	463

Tabla C.57: Tabla de alumnos semana 7 con algoritmo ADTree en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	8030	13
<i>Aprobado</i>	21	472

Tabla C.58: Tabla de alumnos semana 8 con algoritmo ADTree en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	8030	13
<i>Aprobado</i>	21	472

Tabla C.59: Tabla de alumnos semana 9 con algoritmo ADTree en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	8028	15
<i>Aprobado</i>	17	476

Tabla C.60: Tabla de alumnos semana 10 con algoritmo ADTree en MOOC Android

D J48

D.1 Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1743	1
<i>Aprobado</i>	148	51

Tabla D.61: Tabla de alumnos semana 1 con algoritmo J48 en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1743	21
<i>Aprobado</i>	97	102

Tabla D.62: Tabla de alumnos semana 2 con algoritmo J48 en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1736	28
<i>Aprobado</i>	99	100

Tabla D.63: Tabla de alumnos semana 3 con algoritmo J48 en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1755	9
<i>Aprobado</i>	95	104

Tabla D.64: Tabla de alumnos semana 4 con algoritmo J48 en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1758	6
<i>Aprobado</i>	80	119

Tabla D.65: Tabla de alumnos semana 5 con algoritmo J48 en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1743	21
<i>Aprobado</i>	36	163

Tabla D.66: Tabla de alumnos semana 6 con algoritmo J48 en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1747	17
<i>Aprobado</i>	27	172

Tabla D.67: Tabla de alumnos semana 7 con algoritmo J48 en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1745	19
<i>Aprobado</i>	17	182

Tabla D.68: Tabla de alumnos semana 8 con algoritmo J48 en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1758	6
<i>Aprobado</i>	9	190

Tabla D.69: Tabla de alumnos semana 9 con algoritmo J48 en MOOC Don Quijote

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	1758	6
<i>Aprobado</i>	6	193

Tabla D.70: Tabla de alumnos semana 10 con algoritmo J48 en MOOC Don Quijote

D.2 Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	8000	43
<i>Aprobado</i>	385	108

Tabla D.71: Tabla de alumnos semana 1 con algoritmo J48 en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	7995	48
<i>Aprobado</i>	315	178

Tabla D.72: Tabla de alumnos semana 2 con algoritmo J48 en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	7971	72
<i>Aprobado</i>	174	319

Tabla D.73: Tabla de alumnos semana 3 con algoritmo J48 en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	7903	140
<i>Aprobado</i>	59	434

Tabla D.74: Tabla de alumnos semana 4 con algoritmo J48 en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	7933	110
<i>Aprobado</i>	40	453

Tabla D.75: Tabla de alumnos semana 5 con algoritmo J48 en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	7984	59
<i>Aprobado</i>	39	454

Tabla D.76: Tabla de alumnos semana 6 con algoritmo J48 en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	8021	22
<i>Aprobado</i>	25	468

Tabla D.77: Tabla de alumnos semana 7 con algoritmo J48 en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	8034	9
<i>Aprobado</i>	12	481

Tabla D.78: Tabla de alumnos semana 8 con algoritmo J48 en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	8034	9
<i>Aprobado</i>	11	482

Tabla D.79: Tabla de alumnos semana 9 con algoritmo J48 en MOOC Android

<i>Alumnos</i>	<i>Suspenso</i>	<i>Aprobado</i>
<i>Suspenso</i>	8034	9
<i>Aprobado</i>	10	483

Tabla D.80: Tabla de alumnos semana 10 con algoritmo J48 en MOOC Android

E Clusters

E.1 Don Quijote

Alumnos	1	2	3	4
Suspenso	1376	301	82	5
Aprobado	66	56	70	7

Tabla E.81: Tabla de alumnos semana 1 separados por clústeres en MOOC Don Quijote

Alumnos	1	2	3	4
Suspenso	1460	238	58	8
Aprobado	62	51	69	17

Tabla E.82: Tabla de alumnos semana 2 separados por clústeres en MOOC Don Quijote

Alumnos	1	2	3	4
Suspenso	1481	215	61	7
Aprobado	56	35	75	33

Tabla E.83: Tabla de alumnos semana 3 separados por clústeres en MOOC Don Quijote

Alumnos	1	4	5
Suspenso	1564	20	180
Aprobado	41	100	58

Tabla E.84: Tabla de alumnos semana 4 separados por clústeres en MOOC Don Quijote

Alumnos	1	4	5
Suspenso	1578	18	168
Aprobado	28	111	60

Tabla E.85: Tabla de alumnos semana 5 separados por clústeres en MOOC Don Quijote

Alumnos	1	2	3	4
Suspenso	1611	130	20	3
Aprobado	17	42	68	72

Tabla E.86: Tabla de alumnos semana 6 separados por clústeres en MOOC Don Quijote

Alumnos	1	2	3	4
Suspenso	1602	134	22	6
Aprobado	8	16	98	77

Tabla E.87: Tabla de alumnos semana 7 separados por clústeres en MOOC Don Quijote

Alumnos	1	2	3	4
Suspenso	1603	130	22	9
Aprobado	6	12	70	111

Tabla E.88: Tabla de alumnos semana 8 separados por clústeres en MOOC Don Quijote

<i>Alumnos</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
<i>Suspenso</i>	1587	144	23	10
<i>Aprobado</i>	2	5	76	116

Tabla E.89: Tabla de alumnos semana 9 separados por clústeres en MOOC Don Quijote

<i>Alumnos</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
<i>Suspenso</i>	1587	141	25	11
<i>Aprobado</i>	2	4	72	121

Tabla E.90: Tabla de alumnos semana 10 separados por clústeres en MOOC Don Quijote

E.2 Android

<i>Alumnos</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
<i>Suspenso</i>	5389	1882	646	126
<i>Aprobado</i>	66	119	213	95

Tabla E.91: Tabla de alumnos semana 1 separados por clústeres en MOOC Android

<i>Alumnos</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
<i>Suspenso</i>	5667	1765	390	221
<i>Aprobado</i>	34	68	147	244

Tabla E.92: Tabla de alumnos semana 2 separados por clústeres en MOOC Android

<i>Alumnos</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
<i>Suspenso</i>	5692	1649	488	214
<i>Aprobado</i>	7	19	152	315

Tabla E.93: Tabla de alumnos semana 3 separados por clústeres en MOOC Android

<i>Alumnos</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
<i>Suspenso</i>	5726	1589	555	173
<i>Aprobado</i>	1	3	121	368

Tabla E.94: Tabla de alumnos semana 4 separados por clústeres en MOOC Android

<i>Alumnos</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
<i>Suspenso</i>	5796	1563	567	117
<i>Aprobado</i>	0	1	90	402

Tabla E.95: Tabla de alumnos semana 5 separados por clústeres en MOOC Android

<i>Alumnos</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
<i>Suspenso</i>	5918	1502	562	61
<i>Aprobado</i>	0	2	84	407

Tabla E.196: Tabla de alumnos semana 6 separados por clústeres en MOOC Android

<i>Alumnos</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
<i>Suspenso</i>	6065	1416	538	24
<i>Aprobado</i>	0	1	49	443

Tabla E.97: Tabla de alumnos semana 7 separados por clústeres en MOOC Android

<i>Alumnos</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
<i>Suspenso</i>	6176	1350	497	20
<i>Aprobado</i>	0	1	25	467

Tabla E.98: Tabla de alumnos semana 8 separados por clústeres en MOOC Android

<i>Alumnos</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
<i>Suspenso</i>	6184	1339	500	20
<i>Aprobado</i>	0	1	25	467

Tabla E.99: Tabla de alumnos semana 9 separados por clústeres en MOOC Android

<i>Alumnos</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
<i>Suspenso</i>	6185	1338	500	20
<i>Aprobado</i>	0	1	24	468

Tabla E.100: Tabla de alumnos semana 10 separados por clústeres en MOOC Android

